



Escuela
Politécnica
Superior

Exploración de un sistema de detección temprana de las epidemias de gripe a partir de datos climáticos



Grado en Ingeniería Robótica

Trabajo Fin de Grado

Autor:

Enrique Morales Castelló

Tutor/es:

David García García



Universitat d'Alacant
Universidad de Alicante

Exploración de un sistema de detección temprana de las epidemias de gripe a partir de datos climáticos

Autor

Enrique Morales Castelló

Tutor/es

David García García

Dpto. Matemática Aplicada



Grado en Ingeniería Robótica



Escuela
Politécnica
Superior



Universitat d'Alacant
Universidad de Alicante

ALICANTE, Enero 2021

Preámbulo

La razón principal que me ha llevado a la realización de este trabajo fue mi creciente interés en el ámbito de la epidemiología. En este terreno desconocido para mí he podido aplicar conocimientos adquiridos en la carrera englobados dentro de la ciencia de datos e inteligencia artificial. Hasta donde yo sé no es un campo en el que suelen aplicar dichos conocimientos. Sin embargo, las posibilidades a priori parecen prometedoras.

Agradecimientos

La realización de este trabajo no habría sido posible sin la visión y apoyo de mi tutor David García, él me enseñó: la fascinación que produce la exploración de lo desconocido, usar el conocimiento para la búsqueda de soluciones que nos benefician a todos y la superación personal mediante el esfuerzo acompañado de la fe en uno mismo.

Todos los conocimientos adquiridos para la realización de este trabajo no hubieran sido posibles sin la Universidad de Alicante y la Escuela Politécnica Superior, a ellos y a todo su personal les agradezco el tiempo invertido en mi educación y en hacer más comfortable mi estancia dentro del centro.

Me gustaría también agradecer a mi familia, en especial a mi tía y madrina Magdalena, por creer siempre en mí y no darme nunca por perdido. A mi pareja Natalia, por darme la luz y amor que me ayudan a crecer como persona. A mis amigos Héctor y Alexis, por todas esas tertulias nocturnas que hacen reflexionar. Finalmente, a mis compañeros de la segunda promoción, por hacer de este viaje toda una aventura que jamás olvidaré.

It is possible to commit no mistakes and still lose. That is not a weakness; that is life.

Jean-Luc Picard.

Índice general

1. Introducción	1
2. Marco Teórico	3
2.1. Epidemias	3
2.1.1. Gripe	4
2.2. Modelo SIR	6
2.2.1. Obtención de las ecuaciones diferenciales	6
2.2.2. Número básico de reproducción	9
2.3. Redes neuronales	11
2.3.1. Redes neuronales recurrentes RNN	12
2.3.2. Ampliando la memoria de RNN usando LSTM y GRU	12
2.4. Keras	13
3. Objetivos	15
4. Preparación y Visualización de Datos	17
4.1. Obtención del conjunto de datos	17
4.2. Lectura del conjunto de datos	18
4.3. Parámetros y unidades de medida	19
5. Exploración y Planificación del Modelo	23
5.1. Estacionalidad	23
5.1.1. Variables de entrada	24
5.1.2. Variables de salida	31
5.2. Matriz de Correlaciones	33

6. Metodología del Modelo y Resultados	35
6.1. Modelo predictivo de los infectados por gripe	36
6.1.1. Construcción del modelo	36
6.1.2. Experimentos y resultados	38
6.2. Modelo predictivo de la variable beta	41
6.2.1. Resolución del sistema de ecuaciones diferencial SIR	41
6.2.2. Obtención de la variable Beta β y Gamma γ	42
6.2.3. Construcción del modelo	43
6.2.4. Experimentos y resultados	45
7. Conclusiones	49
8. Proyectos futuros	51
Bibliografía	53
Lista de Acrónimos y Abreviaturas	55
A. Anexo I: Gráficas de visualización de las variables del conjunto de datos	57
B. Anexo II: Gráficas de resultados	67
B.1. Resultados modelo predictivo de infectados por gripe	68
B.2. Resultados modelo predictivo de variable beta	71

Índice de figuras

2.1. "Virus de la Viruela." imagen coloreada por Miguel Ángel García licencia bajo BY CC 1.0.	3
2.2. Teniendo a la población en tanto por uno, con las condiciones iniciales establecidas y considerando β y γ constantes. Podemos observar que la evolución de los tres grupos es consistente con nuestra observación del modelo SIR. . .	8
2.3. Número Efectivo de Reproducción para el COVID-19 en España. Cortesía de David García, <i>Dpt. Matemática Aplicada UA</i>	9
2.4. "Estructura básica de una red neuronal convencional (<i>artificial neural network</i>) con una capa oculta." por Colin M.L. Burnett licencia bajo BY CC 3.0. . . .	11
4.1. Visualización numérica mediante una tabla de las variables para las diez primeras semanas de la temporada 2001/2002. Descripción de las unidades en la Sección 4.3	18
4.2. Temporadas de Gripe. Media móvil ponderada a partir de 5 semanas.	21
5.1. Diferentes partes de un diagrama de cajas y bigotes	23
5.2. Diagrama de cajas para la temperatura. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	24
5.3. Diagrama de cajas para la humedad. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	25
5.4. Diagrama de cajas para la temperatura de rocío. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	26
5.5. Diagrama de cajas para el porcentaje de nubes. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	27

5.6. Diagrama de cajas para la radiación neta en superficie. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	28
5.7. Diagrama de cajas para el NAO. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	29
5.8. Diagrama de cajas para el fenómeno atmosférico <i>El Niño</i> . La primera semana del segundo año de la temporada empieza a partir de la semana 14.	30
5.9. Diagrama de cajas para las temporadas de gripe. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	31
5.10. Diagrama de cajas para la variable β del modelo SIR. La primera semana del segundo año de la temporada empieza a partir de la semana 14.	32
5.11. Matriz de correlaciones de las diferentes variables del conjunto de datos. . . .	33
6.1. Primera arquitectura diseñada para la predicción de la epidemia. Experimento 1.	37
6.2. Estructura de las últimas arquitecturas diseñadas para la predicción de la epidemia. Experimento 2.	37
6.3. Combinación de las arquitecturas diseñadas para la predicción de la epidemia. Experimento 3.	38
6.4. Resultados de test sin el uso de la variable β . Leyenda: " <i>true</i> " para los datos reales, " <i>pred</i> " para los datos predichos por la red.	39
6.5. Inicios de la epidemia sin el uso de la variable β	40
6.6. Buscando el valor de γ adecuado.	43
6.7. Primera arquitectura diseñada para la predicción de β . Experimento 4	43
6.8. Estructura de las últimas arquitecturas diseñadas para la predicción de β . Experimento 5.	44
6.9. Combinación de las arquitecturas diseñadas para la predicción de β . Experimento 6.	44
6.10. Resultados de test con el uso de la variable $\beta(t)$	46
6.11. Inicio, final y duración para el Experimento 4	47
6.12. Inicio, final y duración para el Experimento 5	47
6.13. Inicio, final y duración para el Experimento 6	48

6.14. Variable $\beta(t)$. Conjunto de validación. Leyenda: " <i>true</i> " datos reales, " <i>pred</i> " predicciones de la red, " <i>moving mean_true</i> " media móvil de los datos reales y " <i>moving mean_pred</i> " media móvil de las predicciones de la red.	48
A.1. Temperatura asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	58
A.2. Humedad asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	59
A.3. Punto de rocío asociado a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	60
A.4. Porcentaje de nubes asociado a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	61
A.5. Radiación neta en superficie asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	62
A.6. NAO asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	63
A.7. El fenómeno de El Niño asociado a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.	64
A.8. Temporadas de Gripe. Media móvil ponderada a partir de 5 semanas.	65
A.9. La variable beta del modelo SIR, objetivo principal del proyecto. Media móvil ponderada a partir de 5 semanas.	66
B.1. Resultados de conjunto de entrenamiento. Modelo sin cálculo de $\beta(t)$. Experimento 1	68
B.2. Resultados de conjunto de entrenamiento. Modelo sin cálculo de $\beta(t)$. Experimento 2	69
B.3. Resultados de conjunto de entrenamiento. Modelo sin cálculo de $\beta(t)$. Experimento 3	70
B.4. Resultados de conjunto de entrenamiento. Modelo con cálculo de $\beta(t)$. Experimento 4.	71

B.5. Resultados de conjunto de entrenamiento. Aplicando $\beta(t)$ calculada. Experimento 4. Leyenda: "true" datos reales, "pred" predicciones de la red y "true_recalculate" datos reproducidos a partir de la obtención de $\beta(t)$ mediante el sistema SIR.	72
B.6. Resultados de conjunto de entrenamiento. Modelo con cálculo de $\beta(t)$. Experimento 5.	73
B.7. Resultados de conjunto de entrenamiento. Aplicando $\beta(t)$ calculada. Experimento 5. Leyenda: "true" datos reales, "pred" predicciones de la red y "true_recalculate" datos reproducidos a partir de la obtención de $\beta(t)$ mediante el sistema SIR.	74
B.8. Resultados de conjunto de entrenamiento. Modelo con cálculo de $\beta(t)$. Experimento 6.	75
B.9. Resultados de conjunto de entrenamiento. Aplicando $\beta(t)$ calculada. Experimento 6. Leyenda: "true" datos reales, "pred" predicciones de la red y "true_recalculate" datos reproducidos a partir de la obtención de $\beta(t)$ mediante el sistema SIR.	76

Índice de tablas

6.1. Tabla de resultados resumido para la primera fase del proyecto sin el uso de la variable beta.	39
6.2. Tabla de resultados resumido para la segunda fase del proyecto con el uso de la variable beta.	45
6.3. Tabla de Inicio, Final y Duración para los resultados del conjunto de validación o test con la variable β	46

1. Introducción

A lo largo de la historia de la humanidad se han sucedido diferentes epidemias en la sociedad, muchas de ellas causaron estragos en la población debido a los escasos conocimientos sobre virología. Con el tiempo hemos aprendido más sobre el mundo microscópico escondido ante nuestros ojos, hemos realizado clasificaciones de los diferentes agentes genómicos y también hemos comprendido como actúan dentro de nuestro cuerpo. Si bien el estudio de este trabajo no abarca la acción que estos agentes realizan sobre nuestro cuerpo, sí que considera otros aspectos importantes dentro de la epidemiología que son relativos a la manera de expansión que tienen dentro de nuestra sociedad.

Mediante las matemáticas podemos crear modelos que nos permitan comprender la dinámica de estas epidemias. Uno de los modelos básicos y al mismo tiempo eficaz, es el modelo Susceptible Infectado Recuperado (SIR). Este modelo será de principal interés para la realización de este proyecto. Utilizaremos las herramientas actuales dentro del ámbito del Machine Learning (ML) para dar solución a una problemática dentro de las ecuaciones diferenciales que modelan el comportamiento de la epidemia, usando el modelo SIR. Este modelo contiene parámetros que evolucionan dentro del tiempo y actualmente no se conoce con exactitud qué factores influyen en ellos con mayor gravedad, lo que sí se conoce es que algunos de ellos están relacionados con la climatología; por ello, usaremos variables como la temperatura y la humedad, como entradas de nuestro modelo computacional durante su entrenamiento, para tratar de determinar su grado de influencia y obtener un modelo que finalmente produzca predicciones con la mayor exactitud posible; o en su defecto, que nos permita comprender mejor el problema.

El modelo computacional elegido para llevar a cabo la tarea será una *red neuronal recurrente* ya que la evolución de la epidemia a lo largo del tiempo puede ser entendida como una serie temporal, siendo este tipo de red la más aconsejable para modelar este tipo de series. Los datos están medidos a lo largo de intervalos iguales, por semanas. Debemos entrenar a la red para que sea capaz de interpretar la relación que existe entre los datos climáticos y el desarrollo de la epidemia.

La epidemia que se estudiará en este trabajo será la gripe, ya que todos los años se producen brotes en España y disponemos de datos recogidos a partir del Sistema de Vigilancia de la Gripe en España (SVGE)¹. Si establecemos una relación de las variables climáticas y su influencia con respecto a las epidemias, podríamos usar las predicciones climáticas para arrojar luz sobre la dinámica de las epidemias. Esto permitiría ajustar mejor las épocas de vacunación, gestionar recursos materiales y humanos en el pico de la epidemia, y reducir, en definitiva, el impacto de la gripe en España, el cual está valorado en millones de euros cada año (Pérez-Rubio Alberto, 2019).

¹<https://vgripe.isciii.es>

2. Marco Teórico

2.1. Epidemias

Corría el año 1521; Hernán Cortés, al frente de su ejército, conquistaba la gran capital Tenochtitlan, poniendo fin a dos siglos de esplendor del imperio Azteca. El conquistador, junto con los demás hombres que lo acompañaban, fueron reconocidos como héroes nacionales al ser capaces de someter a un imperio con los escasos efectivos y medios de los que disponían.

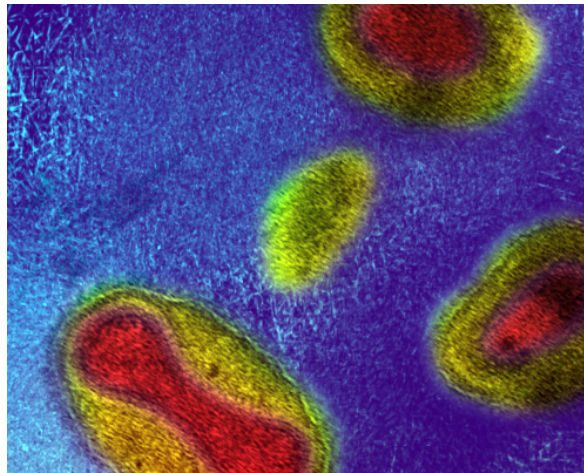


Figura 2.1: "Virus de la Viruela." imagen coloreada por Miguel Ángel García licencia bajo BY CC 1.0.

No obstante, la historia de la conquista de América por los europeos tiene un protagonista olvidado: la *viruela*; este virus que portaban los Españoles se esparció rápidamente entre la población indígena, mermando a sus ejércitos y debilitando a sus ciudadanos. El historiador y demógrafo Noble David Cook estima que las regiones menos afectadas perdieron un 80% de su población y las más afectadas entre un 90% y 100% (Cook, 1998).

Como se puede observar, la historia contada en la introducción de este capítulo es un claro ejemplo de lo importante que pueden ser las epidemias en el transcurso de nuestra historia. Una pandemia, como la que actualmente sufrimos producida por la COVID-19, nos hace cambiar hábitos, nos recluye en nuestras casas, hunde la economía de países enteros o colapsa nuestros hospitales, que no dan abasto ante la avalancha de afectados.

Entender cómo se propaga un virus en la sociedad es de vital importancia y desarrollar métodos de predicción puede ayudarnos a prevenir futuros brotes. No obstante, esta tarea no es fácil debido a que hay diversos factores que pueden afectar, tales como: casos importados, costumbres locales, población vacunada, entre muchos otros...

2.1.1. Gripe

El virus de la gripe es tan antiguo como la historia de la humanidad, disponemos de escritos provenientes del año 412 A.C que describen posibles brotes epidémicos, con ello, a lo largo de los tiempos se han sucedido diferentes pandemias relacionadas con este virus. Aún se recuerda la acaecida en el año 1918, con un rebrote en 1919, conocida como la *gripe española*, llamada así porque los medios Españoles no censuraron información sobre la misma, tal y como sucedió con los países que estaban involucrados en la I guerra mundial, ya que no querían dar información extra a sus enemigos. La pandemia mencionada golpeo al mundo en tres oleadas diferentes, siendo la segunda la que causó más muertes en cuestión de horas o días una vez desarrollados los síntomas. La gripe española tiene gran trascendencia en los acontecimientos que estaban por suceder a principios del siglo XX. Recientemente se ha estudiado de forma más concreta los acaecidos en la Alemania de ese periodo, donde se ha demostrado que existe una correlación de las muertes provocadas por el virus con el ascenso al poder de partidos políticos que trazaron una senda que desembocó en la segunda gran guerra mundial (Blickle, 2020).

Este tipo de virus, clasificado como tipo ARN, se puede entender como una máquina replicadora, tal y como lo describe John von Neumann en su libro *The Theory of Self-reproducing Automata*. En él habla de las bases de la autoreplicación, donde compone a un sistema de tres

partes etiquetadas como A (constructor), B (duplicador) y C (controlador). Para que una máquina sea capaz de replicarse a sí misma necesita una serie de instrucciones (A, B, C) y esta lógica también puede ser aplicada a las células. Hoy sabemos que un virus como la gripe utiliza las instrucciones A y B de una célula de nuestro organismo para generar copias de sí mismo.

En España, las temporadas de gripe duran de octubre a mayo. El Sistema de Vigilancia de la Gripe en España recoge datos de dichas temporadas. También se sabe que la capacidad de propagación y supervivencia del virus está correlacionada con factores climáticos, entre los cuales podemos destacar (Fuhrmann, 2010):

- **Temperatura:** Siendo las temporadas con temperaturas más bajas las facilitan la supervivencia del virus.
- **Humedad:** La sequedad del ambiente hace que el virus viaje más fácilmente por medio de aerosoles.
- **Radiación solar:** Temporadas de baja radiación ultravioleta (UV) reducen los niveles de vitamina D, lo cual perjudica al sistema de péptidos antimicrobianos del cuerpo, responsable de regular la respuesta inmune (Cannel et al., 2006, 2008).

Realizamos especial hincapié en la humedad del ambiente, ya que ambientes más secos permiten que las gotas salivares se hagan más pequeñas por evaporación al ser expulsadas. Cuando éstas son de un tamaño inferior a $5\ \mu m$ permanecen suspendidas en el aire durante más tiempo, y esto hace que el virus pueda viajar a mayor distancia entre el huésped y un susceptible. El medio aéreo es uno de los principales modos de transmisión viral de este virus, siendo otros:

Contacto físico: las gotas salivares caen directamente sobre el susceptible.

Fomites: objetos inanimados que han entrado en contacto con el agente infeccioso, la mayoría de ellos comunes.

2.2. Modelo SIR

Este modelo está elaborado dividiendo a una población de individuos en tres estados diferentes, los cuales dan nombre al mismo, que utilizaremos en la realización de este trabajo.

1. **Susceptibles**, $S(t)$: Pueden contraer la enfermedad y aún no han sido infectados por ella.
2. **Infectados**, $I(t)$: Los que han contraído la enfermedad y pueden contagiarla a los susceptibles.
3. **Recuperados**, $R(t)$: Individuos que contrajeron la enfermedad pero que ya no la transmiten como los Infectados. Ya sea porque han sanado o fallecido.

2.2.1. Obtención de las ecuaciones diferenciales

El modelo SIR nos dice cómo evoluciona el número de individuos que se encuentran en cada uno de los tres grupos señalados. Para construirlo, primero, debemos de establecer una serie de condiciones, las cuales son:

- El número de individuos en nuestra población total es siempre constante, queda expresado por N .
- Los infectados pueden transmitir a varios susceptibles la enfermedad. Usaremos un parámetro β , que representa el promedio de contagios por unidad de tiempo, y nos dará la información sobre la capacidad de transmisión del virus por parte de este grupo.
- La capacidad de recuperación de los infectados queda determinada por un parámetro γ , determinado como la inversa de la duración de la infección.

Una vez establecidas las condiciones, podemos formular la primera expresión de nuestro modelo, que expresa que la suma de los tres grupos mencionados es igual a la población total:

$$S + I + R = N \tag{2.1}$$

Seguidamente, pasamos a formular las diferentes derivadas de las funciones que representan la evolución del número de individuos en cada grupo. En primer lugar tenemos el grupo de recuperados $R(t)$, que depende directamente de $I(t)$, ya que una disminución de los infectados aumentará el grupo de los recuperados, ningún individuo de este último grupo volverá al grupo de infectados¹. La velocidad de trasvase de $I(t)$ a $R(t)$ está controlada por el parámetro γ de forma que,

$$\frac{dR}{dt} = \gamma I \quad (2.2)$$

Los susceptibles $S(t)$ decrecen en el tiempo, a un ritmo controlado por el parámetro β y proporcional al número de contactos entre S e I, el cual se representa como el producto $S \cdot I$. La siguiente expresión en nuestro modelo será la derivada de $S(t)$, que será negativa al ser decreciente y proporcional al producto señalado de $S \cdot I$, de forma que,

$$\frac{dS}{dt} = -\beta SI \quad (2.3)$$

La última expresión, la que nos indica la evolución de los infectados $I(t)$, queda determinada a partir de las dos expresiones anteriores y de la expresión 2.1. Los infectados aumentarán tanto como disminuyan los susceptibles y decrecerán tanto como aumenten los recuperados, de forma que,

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (2.4)$$

¹Consideramos que una vez recuperado, el individuo es inmune al virus.

En resumen, el modelo SIR viene dado por el sistema de ecuaciones diferenciales:

$$\frac{dR}{dt} = \gamma I \quad (2.2)$$

$$\frac{dS}{dt} = -\beta SI \quad (2.3)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (2.4)$$

Podemos utilizar MATLAB® para obtener una aproximación de las soluciones de estas ecuaciones. Para los parámetros $\beta = 0.49$, $\gamma = 0.14$ y $N = 1$, y las condiciones iniciales $S(0) = 1$, $I(0) = 0.01$ y $R(0) = 0$. Tomando $N = 1$ se entiende que los valores de S , I y R son proporcionales respecto al total de la población. La Figura 2.2 muestra como los susceptibles siempre decrecen, mientras que los recuperados se incrementan. Por otro lado, los infectados tienen una primera fase de crecimiento exponencial seguida de una desaceleración del crecimiento, un máximo y un decrecimiento. Nótese que los susceptibles se estabilizan en 0.034, es decir, no todos se contagian. Esta asíntota es conocida como la *inmunidad de rebaño*.

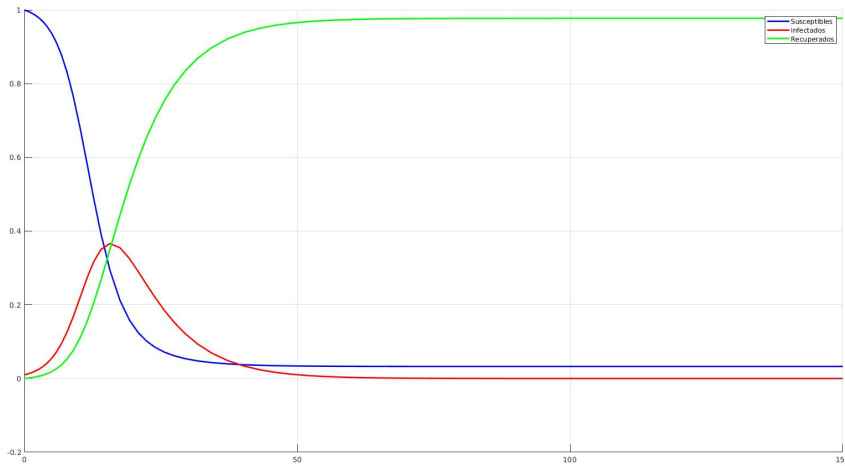


Figura 2.2: Teniendo a la población en tanto por uno, con las condiciones iniciales establecidas y considerando β y γ constantes. Podemos observar que la evolución de los tres grupos es consistente con nuestra observación del modelo SIR.

2.2.2. Número básico de reproducción

Este número es el cociente entre el parámetro β y γ , $R = \frac{\beta}{\gamma}$.

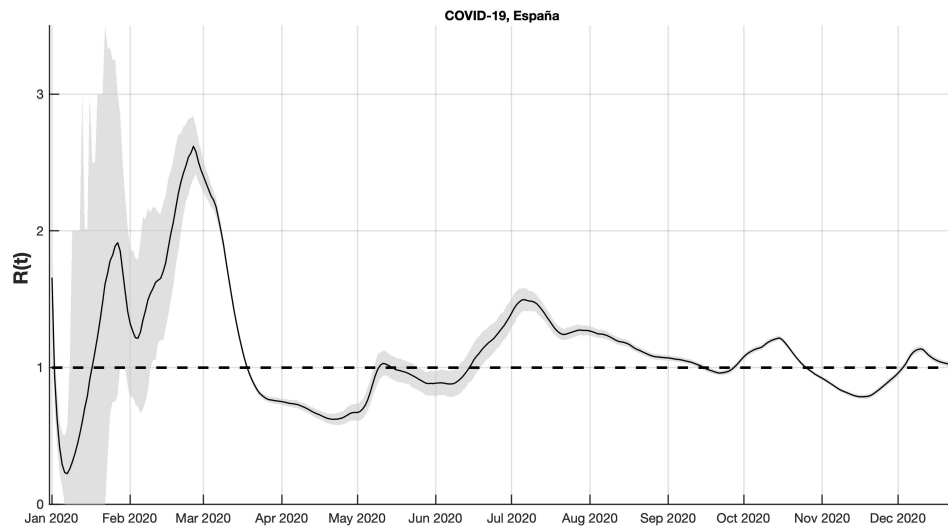


Figura 2.3: Número Efectivo de Reproducción para el COVID-19 en España. Cortesía de David García, *Dpt. Matemática Aplicada UA*

Nos dice cuántos susceptibles son contagiados en promedio por un único infectado y también nos informará de la aparición de un brote si éste es mayor que 1.

El Número Básico de Reproducción (NBR) es el que se calcula para el primer infectado, por eso se escribe R_0 . A partir de este momento lo que se calcula es el Número Efectivo de Reproducción (NER), y se representa con $R(t)$. Es un dato importante a tener en cuenta ya que nos puede dar información útil, por ejemplo, sobre si las medidas adoptadas por una sociedad son buenas para frenar una epidemia.

En la Figura 2.3 podemos observar como hasta mediados de marzo, con las medidas tomadas en el confinamiento por el gobierno, se reduce por debajo de 1. A mediados de junio, conforme se relajan dichas medidas, el valor se sitúa por encima de 1 hasta finales de octubre hasta que se toman nuevamente medidas más restrictivas.

Es importante señalar que en la Figura 2.2 hemos considerado β y γ como constantes. γ se puede considerar constante, pero la situación se complica cuando β evoluciona en el tiempo. La Figura 2.3 muestra la evolución de $R(t)$ producida por cambios en β , es decir: $R = \frac{\beta(t)}{\gamma}$.

Recordemos que este parámetro representa el promedio de contagios por unidad de tiempo, que además está relacionado con la probabilidad de contagio. El distanciamiento entre la ciudadanía durante el confinamiento, así como las medidas higiénicas adoptadas, influyó en la disminución del parámetro $\beta(t)$. **En el caso de la gripe, el parámetro β puede variar debido a cambios en las condiciones climáticas.** Llegados a este punto nos enfrentamos a la gran cuestión que humildemente vamos a intentar desgranar en este proyecto: ¿podemos determinar la variación temporal de $\beta(t)$ teniendo en cuenta datos externos de condiciones atmosféricas como temperatura o humedad?

2.3. Redes neuronales

Basadas en un modelo computacional dentro del área del ML, en el que tenemos unidades básicas, conocidas como neuronas, conectadas entre sí y cuya salida está condicionada por una función de activación. La salida de estas redes se expresa mediante una función que trataremos de minimizar modificando los valores de los pesos de las neuronas usando un algoritmo conocido como *backpropagation*. Son la piedra angular del Deep Learning (DL), siendo una herramienta versátil capaz de resolver tareas de clasificación, entre otros.

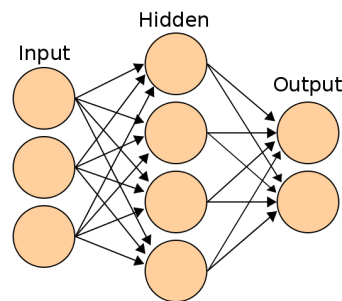


Figura 2.4: "Estructura básica de una red neuronal convencional (*artificial neural network*) con una capa oculta." por Colin M.L. Burnett licencia bajo BY CC 3.0.

Actualmente, gracias al incremento de la potencia computacional desde la década de los 90, ahora es posible entrenar redes más densas (Deep Neural Network (DNN)) en un periodo de tiempo razonable; esto quiere decir, con muchas más neuronas dentro de lo que se conoce como *capas ocultas*. La posibilidad en estos tiempos de incluir más capas ocultas, junto con sus correspondientes neuronas permite resolver algunos problemas que surgían en el pasado con las redes neuronales convencionales, como por ejemplo, la resolución de la puerta lógica *XOR*. La industria del *gaming* también ha ayudado al desarrollo de las mismas ya que potenció la producción de tarjetas gráficas equipadas con GPU, siendo útiles a la hora de aprovechar la potencia computacional que éstas les brindan.

Actualmente la rama del *Machine Learning* conocida como *Deep Learning*, goza de gran popularidad dentro de la inteligencia artificial. Podemos ver su impacto en los primeros coches autónomos, en tareas de reconocimiento de patrones o clasificación de imágenes.

2.3.1. Redes neuronales recurrentes RNN

Recurrent Neural Network (RNN), este tipo de redes se utilizan principalmente para predecir futuras salidas de una secuencia dado un entrenamiento previo con diferentes secuencias de carácter similar. Podemos ver su uso en meteorología, para la predicción del clima o en la conducción autónoma, tratando de anticipar una trayectoria para evitar accidentes. No obstante, otro tipo de redes como las Deep Neural Network o Convolutional Neural Network (CNN) también pueden ser usadas, aunque normalmente son famosas por encargarse de otro tipo de tareas. Este modelo contiene un tipo de unidad básica parecida a las neuronas de las redes neuronales convencionales, esta se conoce como *celdas* y son capaces de conservar información a lo largo del tiempo.

A diferencia de las redes neuronales convencionales, que aceptan una entrada y producen una salida, también constan de una entrada que recibe una salida de un tiempo anterior, podemos entender este concepto como una realimentación de la misma red. Podemos, a partir de este tipo de redes, sacar un solo valor a partir de una secuencia o una secuencia entera a partir de otra secuencia, utilizando la salida de la primera como un vector y la entrada de la segunda como un vector, esto último se utiliza normalmente para sistemas que traducen de una lengua a otra.

El problema principal de este tipo de redes es que resulta difícil entrenarlas para que procesen secuencias que son demasiado largas, ya que a la larga en su entrenamiento, olvidarán las primeras entradas de la secuencia.

2.3.2. Ampliando la memoria de RNN usando LSTM y GRU

Tal y como se ha señalado en el último párrafo del apartado anterior, sería conveniente disponer de un modelo que solucionase, en mayor medida, la problemática de poder entrenar la red con secuencias más largas. Para lograrlo, se han creado otros modelos que modifican las unidades de memoria conocidas como celdas de las redes recurrentes básicas. Este otro tipo de modelos o redes son conocidas como Long Short-Term Memory (LSTM) y Gated Recurrent Unit (GRU). Son considerablemente más efectivas que las RNN.

- **LSTM:** La mejora que introducen respecto a las RNN consiste en que son capaces de reconocer la información más relevante de la entrada, siendo capaz de preservarla todo el tiempo que sea necesario. Esta celda contiene tres puertas que regulan el tránsito de información: *forget gate*, decide cuánta información retenemos u olvidamos; *input gate*, actualiza el estado de la celda, esto hace que la red neuronal encuentre valores relevantes; finalmente, *output gate*, que decide cuál será el próximo estado, que contendrá información de entradas previas.
- **GRU:** Puede entenderse como una versión simplificada de la LSTM, donde podemos encontrar dos puertas: *update*, donde se decide qué información se olvida y cual se añade y *reset*, donde se decide cuanta información pasada se puede olvidar.

2.4. Keras

Keras se trata de una Application Programming Interface (API) de alto nivel que nos permite construir, entrenar y evaluar redes neuronales. Fue lanzada como un proyecto de código abierto en 2015, con el tiempo ganó popularidad volviéndose más robusta y eficaz abordando temas complejos. Para este proyecto se ha escogido usar Keras junto con TensorFlow como backend, este último es una biblioteca software desarrollada por Google.

3. Objetivos

Se resume, brevemente, en este apartado los principales objetivos de este proyecto. Así también se señalizan las principales fuentes de información que se usarán para estudiar cada uno de los mismos. En orden cronológico, las fases del proyecto consistirán en:

1. **Recopilación de los datos climáticos:** se realizará a través de los datos de libre acceso proporcionados por el Centro Europeo de Previsiones Meteorológicas a Plazo Medio (European Centre for Medium-Range Weather Forecast (ECMWF), por sus siglas en inglés) y del Sistema de Vigilancia de la Gripe en España.
2. **Preparación de los datos:** se exportarán los datos, utilizando *Matlab R2018b*, a un formato que permita organizar y dar un rápido acceso a los mismos cuando utilicemos, en las fases posteriores, las herramientas necesarias para su análisis.
3. **Python:** familiarizarse con el lenguaje de programación *python*; además, se utilizarán las librerías como *pandas*, *matplotlib*, *numpy*, entre otras, para visualizar y organizar los datos.
4. **Deep Learning:** se realizará un estudio previo de la librería Keras, en concreto, del apartado de *redes neuronales recurrentes* para llevar a cabo los experimentos.
5. **Jupyter Notebook:** usando esta aplicación web comenzaremos a organizar los diferentes bloques del proyecto, también visualizaremos y limpiaremos los datos, dándoles un formato adecuado para la realización de los experimentos.
6. **Primeros prototipos:** en esta fase empezaremos a obtener los primeros resultados con arquitecturas simples que utilizarán directamente los datos climáticos para predecir la evolución de la epidemia. En esta fase no estaremos usando las ecuaciones SIR (sección 2.2).

7. **Cálculo del parámetro beta:** Utilizando un método de aproximación obtendremos la resolución del sistema SIR, a partir del cual obtendremos los valores de $\beta(t)$ que utilizaremos en nuestros prototipos finales.
8. **Prototipos finales:** Se procederá a realizar predicciones utilizando los valores de $\beta(t)$.
9. **Resultados y conclusiones finales:** En esta última fase compararemos los resultados obtenidos con las diferentes arquitecturas utilizadas, así mismo obtendremos las conclusiones finales que podrán ser utilizadas para futuras investigaciones relacionadas con este proyecto.

En base a los objetivos marcados podemos asegurar que la aplicabilidad del proyecto está asegurada, ya que el acceso a los datos es gratuito y su distribución libre. Además, el software que se utilizará, usando librerías de python como *Keras*, también es de libre acceso.

Se espera tener una comprensión más amplia sobre la relación de las variables climáticas con la evolución de las epidemias de gripe. Esta comprensión aspira a ser de ayuda en el ajuste de las épocas de vacunación, gestión de recursos materiales y humanos. Todo ello permitiría reducir el impacto de la epidemia sobre el país.

4. Preparación y Visualización de Datos

En este apartado se discutirá sobre la obtención de los datos y el preprocesamiento de los mismos, esto es: detectar la ausencia de valores, observar si existen inconsistencias por un mal almacenamiento, detectar valores duplicados, outliers o valores atípicos. La última etapa del preprocesado de los datos consistirá en la transformación de los mismos a un formato que permita la manipulación correcta por parte de las diferentes arquitecturas que utilicemos durante el proyecto. Las temporadas están separadas entre sí por las treinta y dos semanas que dura cada una de ellas.

4.1. Obtención del conjunto de datos

- Los datos de *temperatura de la atmósfera, temperatura de rocío, porcentaje de nubes y radiación neta en superficie* provienen del conjunto de datos European Research Area (ERA5)¹, de libre acceso, siendo mallas de paralelos y meridianos de $0.25^0 \times 0.25^0$. Dichos datos han sido promediados en las semanas de interés en la región de España.
- Los datos de infectados por gripe provienen del Instituto de Salud Carlos III, la Red Nacional de Vigilancia Epidemiológica (RENAVE) y SVGE. A partir de ellos y con las ecuaciones del SIR (sección 2.2) hemos obtenido los valores para la variable *beta*.
- Los datos provenientes para North Atlantic Oscillation (NAO) y Fenómeno El Niño (FEN) han sido obtenidos a partir de Koninklijk Nederlands Meteorologisch Instituut (KNMI)².
- Debido al desfase que existe en nuestro calendario gregoriano, hubo que tener especial

¹<https://www.ecmwf.int>

²<https://www.knmi.nl>

cuidado en la obtención de los datos, ya que éstos debían coincidir en el mismo día relacionado con el comienzo de la temporada de gripe para realizar este proyecto. La primera semana de la temporada de gripe empieza todos los años en la semana 40 y finaliza en la 20 del año siguiente, esto es, desde principios de octubre hasta mediados de mayo del año siguiente, unas 33 semanas en total.

4.2. Lectura del conjunto de datos

Previamente leemos el archivo que contiene todos los datos numéricos de las variables, dicho archivo ha sido obtenido usando *Matlab*, organizando cada una de las variables.

	Temperatura	Humedad	Rocío	Nubes	Radiación	NAO	Niño	Gripe	Beta
Semana 1-2001/2002	291.6117	92.2955	286.3180	0.1701	439601.1446	-0.2661	-0.2197	0.0050	1.34
Semana 2-2001/2002	289.5804	92.3547	284.3740	0.1123	410893.8216	0.5220	-0.2000	0.0067	1.60
Semana 3-2001/2002	289.3324	93.4898	284.9200	0.1474	358329.0042	-0.6970	-0.2051	0.0107	1.02
Semana 4-2001/2002	287.9974	93.1067	283.3433	0.1014	395216.7600	-0.6214	-0.1397	0.0109	0.46
Semana 5-2001/2002	288.9796	90.0463	282.1439	0.0277	405030.6813	0.5834	0.0782	0.0051	2.96
Semana 6-2001/2002	283.6148	91.8888	278.1988	0.1539	323816.1516	-0.1499	-0.0086	0.0149	1.13
Semana 7-2001/2002	280.0871	91.1360	274.2190	0.1196	288276.9474	-0.0886	-0.2925	0.0168	1.25
Semana 8-2001/2002	280.5843	93.1053	276.0509	0.0683	302640.4736	0.6914	-0.4630	0.0210	0.91
Semana 9-2001/2002	281.2095	93.6193	277.0174	0.1540	283512.5148	1.4596	-0.4353	0.0191	0.61
Semana 10-2001/2002	281.2803	93.8996	277.2754	0.0775	306779.9830	1.0779	-0.4326	0.0116	1.50

Figura 4.1: Visualización numérica mediante una tabla de las variables para las diez primeras semanas de la temporada 2001/2002. Descripción de las unidades en la Sección 4.3

Tras la lectura del fichero contenedor los datos se organizan en una tabla. La Figura 4.1 muestra el contenido para las diez primeras semanas de la temporada 2001/2002. Todas ellas son variables climáticas de entradas a excepción de las variables que intentará predecir nuestra red, siendo estas: la variable *gripe*, referida a los infectados de la población, usada en los primeros prototipos y *beta* la que se usará en los finales, siendo esta última el objetivo principal de este proyecto. Nos aseguramos de no tener datos nulos que corrompan nuestra fuente.

4.3. Parámetros y unidades de medida

Temperatura: La temperatura general asociada a la semana. Siendo las temporadas con temperaturas más bajas las facilitan la supervivencia del virus.

Unidad de Medida: Kelvin (K)

Humedad: La humedad relativa expresada como un porcentaje, siendo esta la relación entre la cantidad de vapor de agua que contiene el aire y la que necesitaría contener para saturarse a la misma temperatura. En el 100% se alcanza el punto de rocío. Cuanto menor sea la misma en el ambiente, mayor facilidad tendrá la epidemia para expandirse, ya que en ambientes más secos las partículas salivares viajan a mayor distancia.

Unidad de Medida: Porcentaje

Rocío: La temperatura de rocío asociada a la semana. El punto de rocío varía durante todo el año al igual que la temperatura media. Al estar relacionada esta variable con una de las principales de este estudio, consideramos relevante su inclusión dentro del conjunto de datos.

Unidad de Medida: Kelvin (K)

Nubes: Expresada como una porción, siendo 0 un cielo totalmente despejado, y 1 un cielo totalmente cubierto. La porción de cobertura de nubes a lo largo del año es importante ya que está relacionado también con las horas de sol a lo largo del día, que a su vez están relacionadas con la fortaleza del sistema inmune, como se señaló en la Sección 2.1.1.

Unidad de Medida: Adimensional

Radiación: La radiación electromagnética que llega a la atmósfera. Medida en superficie horizontal. La radiación neta en superficie es importante ya que el virus es sensible a la misma, su estudio puede ayudar a esclarecer la importancia de las horas de sol diarias.

Unidad de Medida: (W/m^2)

NAO: Oscilación del Atlántico Norte (NAO), fenómeno climático en el norte del océano Atlántico, ocasiona fluctuaciones climáticas. Lo medimos con el índice NAO, que se

calcula como la diferencia entre la presión atmosférica de las Azores e Islanda. Considerada importante en este proyecto ya que es responsable de la variabilidad del clima, como la temperatura y humedad. En España se ha estudiado la influencia de este fenómeno frente a los regímenes de precipitación mensual, estacional y anual (Castro et al., 2011).

Unidad de Medida: Adimensional

Niño: Fenómeno El Niño, fenómeno climático relacionado con el calentamiento del Pacífico oriental ecuatorial, se manifiesta de forma cíclica en periodos de entre tres u ocho años. Lo medimos con el Southern Oscillation Index (SOI), calculado como la diferencia de presión atmosférica entre Tahiti (Polinesia Francesa) y Darwin (Australia). Los periodos en los que esto ocurre van de entre tres a siete años. Este fenómeno tiene influencia en el clima a escala global (Ropelewski y Halpert, 1987).

Unidad de Medida: Adimensional

Gripe: La variable de infectados, llamada en este dataframe: *gripe*, es la porción de afectados por cada 100000 habitantes. Nuestro conjunto de datos posee las diferentes epidemias acaecidas desde la temporada 2001/2002 hasta 2016/2017, tal y como se puede apreciar en la Figura 4.2. Aunque el pico de epidemia suele concentrarse a mediados de la temporada, también podemos observar la presencia de temporadas anómalas dentro del conjunto de datos, en las cuales la epidemia alcanza antes su pico máximo, como por ejemplo en las temporadas 2003/2004 y 2009/2010.

Unidad de Medida: Adimensional

Beta: Promedio de contagios por unidad de tiempo (en nuestro caso semanas). En la Sección 2.2 podemos encontrar información relevante sobre la importancia de esta variable. Es la variable que aspiramos predecir con nuestro modelo, regula la expansión de la epidemia. Los datos almacenados para esta variable han sido calculados a partir de una resolución aproximada del sistema SIR (vease la Sección 6.2.1).

Unidad de Medida: contagios / semana

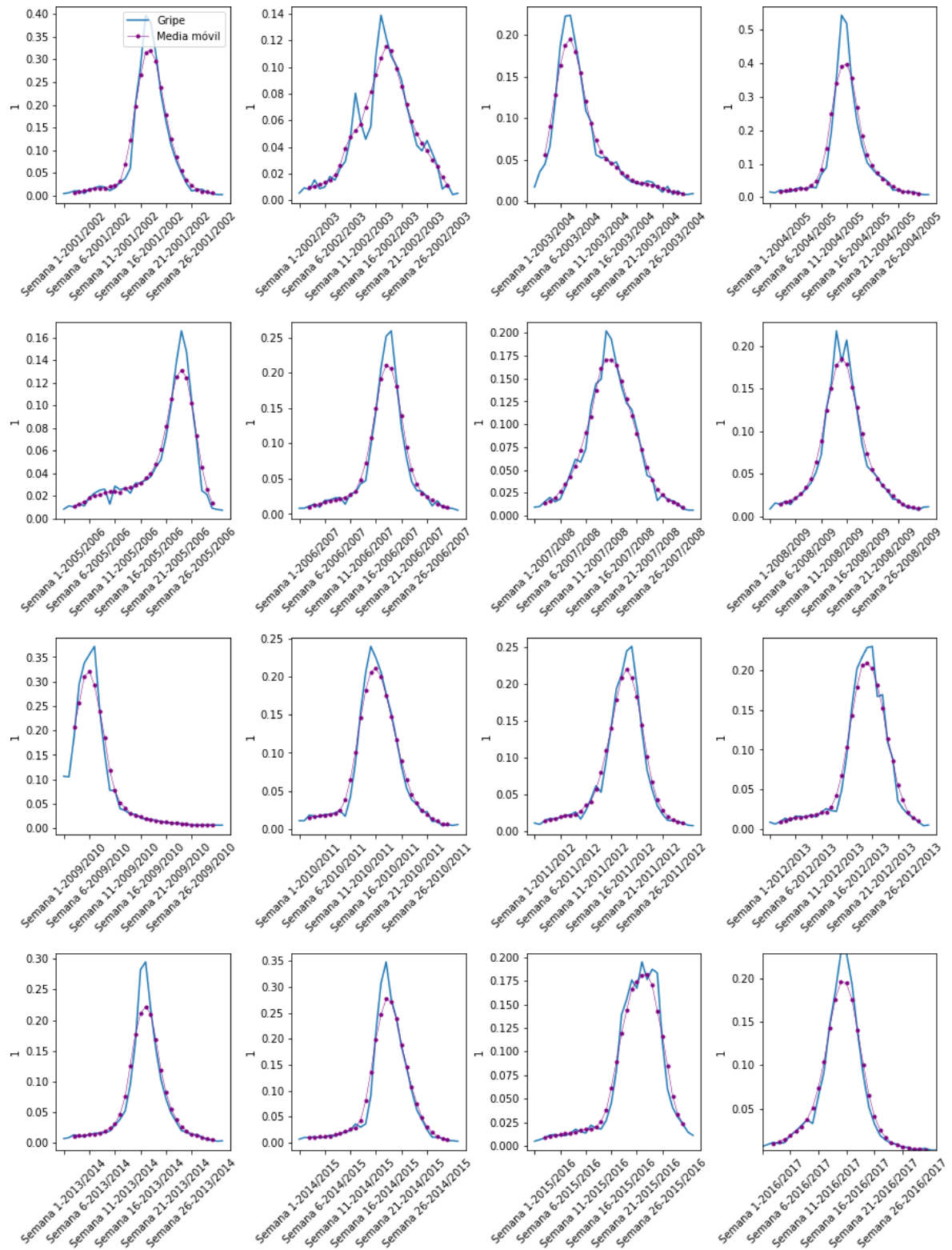


Figura 4.2: Temporadas de Gripe. Media móvil ponderada a partir de 5 semanas.

5. Exploración y Planificación del Modelo

Se realizará un paso previo antes de implementar el modelo que usaremos para realizar las predicciones. En esta fase trataremos de extraer información de valor de nuestros datos, también trataremos de normalizar los mismos, preprocesarlos y darles un formato adecuado para su consolidación dentro de este proyecto.

5.1. Estacionalidad

Vamos a tratar de obtener información relevante analizando la estacionalidad o año promedio a través de *diagramas de cajas*, este diagrama nos permite representar de manera gráfica la distribución de puntuaciones dentro de una variable, esta forma de describir las puntuaciones nos permite ver su grado de asimetría, valores extremos o atípicos, posición de la mediana, entre otros.

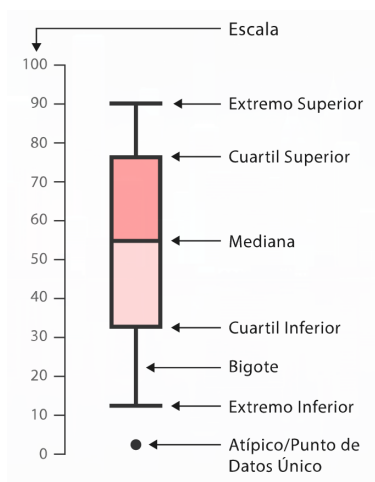


Figura 5.1: Diferentes partes de un diagrama de cajas y bigotes

Como podemos ver en la Figura 5.1, los cuartiles superior e inferior delimitan los valores

de la variable para aquellos que se encuentran dentro de un rango típico del 50%, más allá de ellos encontramos los bigotes y extremos superior e inferior donde se encuentran valores considerablemente altos. Finalmente, los valores atípicos están completamente fuera del diagrama de cajas y se representan normalmente con una señalización en forma de punto o estrella dentro del gráfico.

5.1.1. Variables de entrada

Temperatura: Observamos una clara estacionalidad. Los primeros cuartiles son menores a mitad de temporada, siendo estos los meses más fríos. También observamos la existencia de algunos valores atípicos causados seguramente por condiciones extremas del clima. Durante toda la temporada tenemos que la variabilidad de la temperatura es similar en casi todas las semanas.

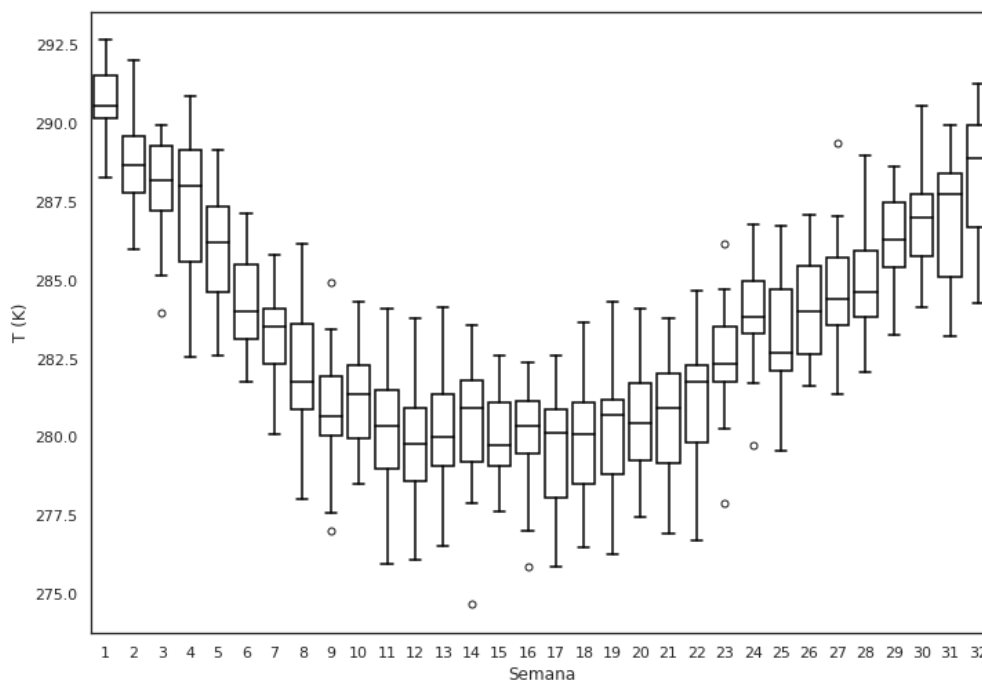


Figura 5.2: Diagrama de cajas para la temperatura. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

Humedad: Observamos una clara estacionalidad. Podemos observar también valores atípicos, sobre todo en las primeras y últimas semanas. Tenemos un crecimiento ascendente en dirección a las semanas más frías de la temporada, seguido de un decrecimiento a partir de la semana 14 correspondiente a la primera semana del segundo año de la temporada.

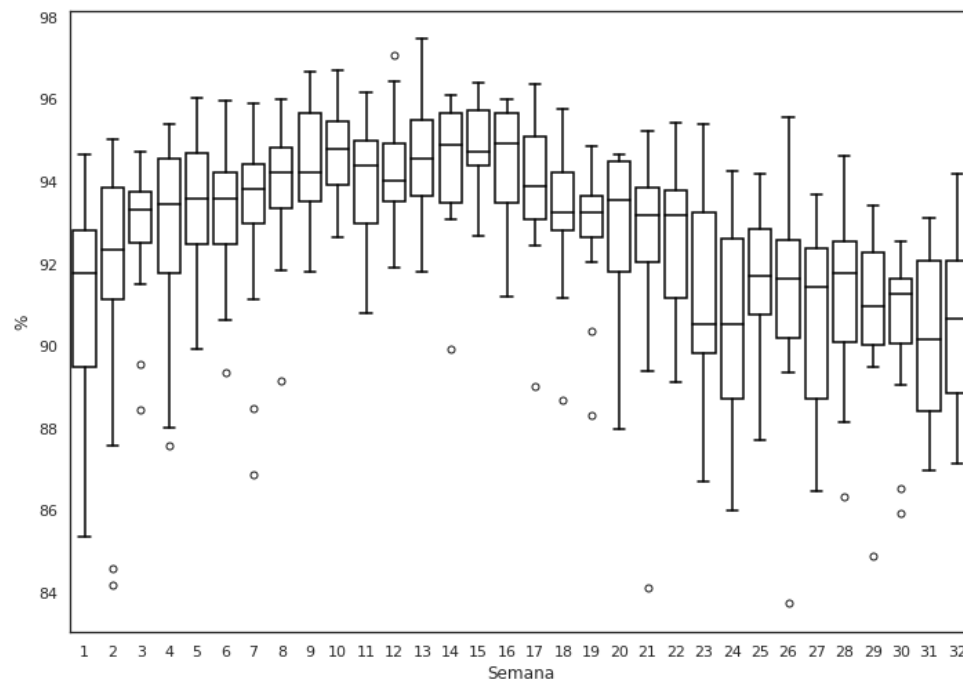


Figura 5.3: Diagrama de cajas para la humedad. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

Rocío: La temperatura de rocío al estar directamente relacionada con la temperatura atmosférica observamos que tenemos un diagrama similar al observado en la Figura 5.2.

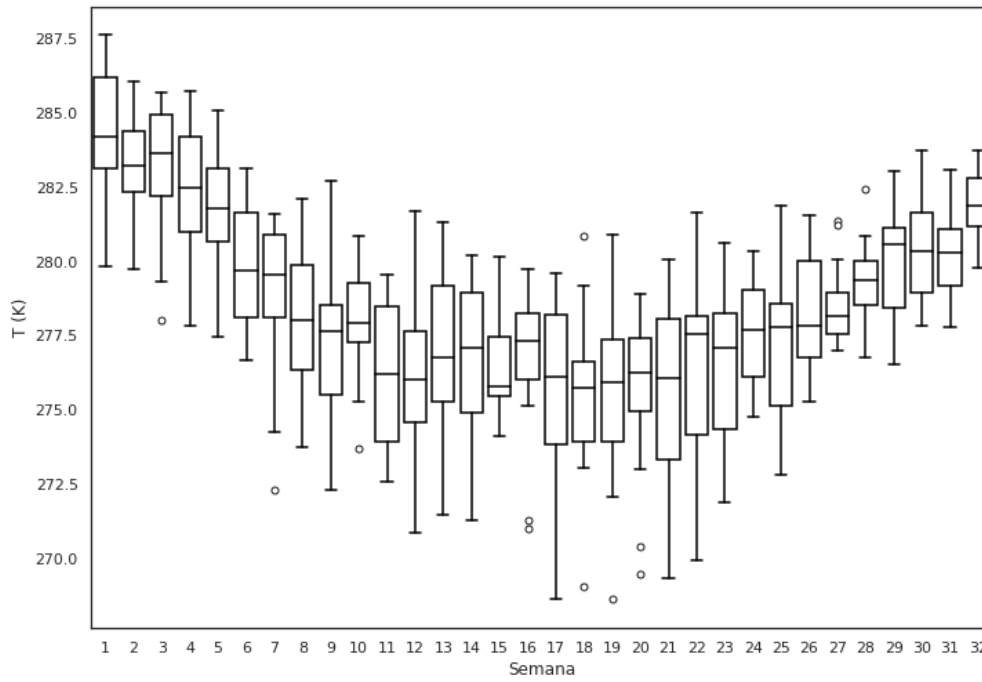


Figura 5.4: Diagrama de cajas para la temperatura de rocío. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

Nubes: La porción de nubes presenta mucha variabilidad para todas las semanas de la temporada. Cuesta encontrar una estacionalidad, aunque los valores del tercer cuartil son mayores a mitad de temporada cuando la humedad también aumenta.

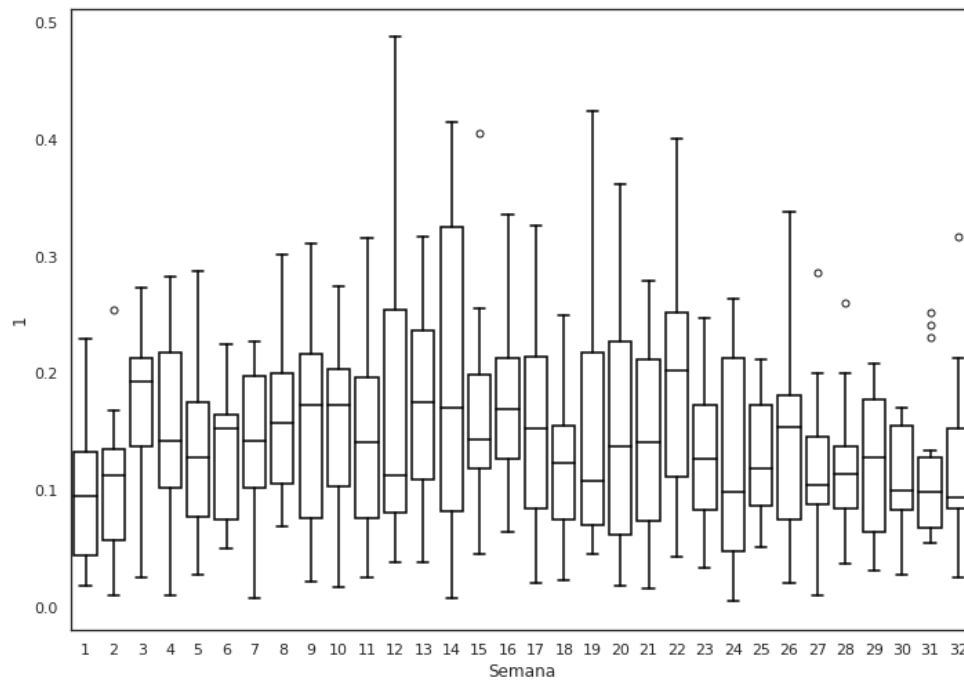


Figura 5.5: Diagrama de cajas para el porcentaje de nubes. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

Raciación: Los valores para la radiación neta en superficie son interesantes ya que podemos observar que no presentan gran variabilidad de una temporada a otra, teniendo una tendencia y estacionalidad claras. A mitad de temporada podemos observar que los valores para la mediana son los más bajos.

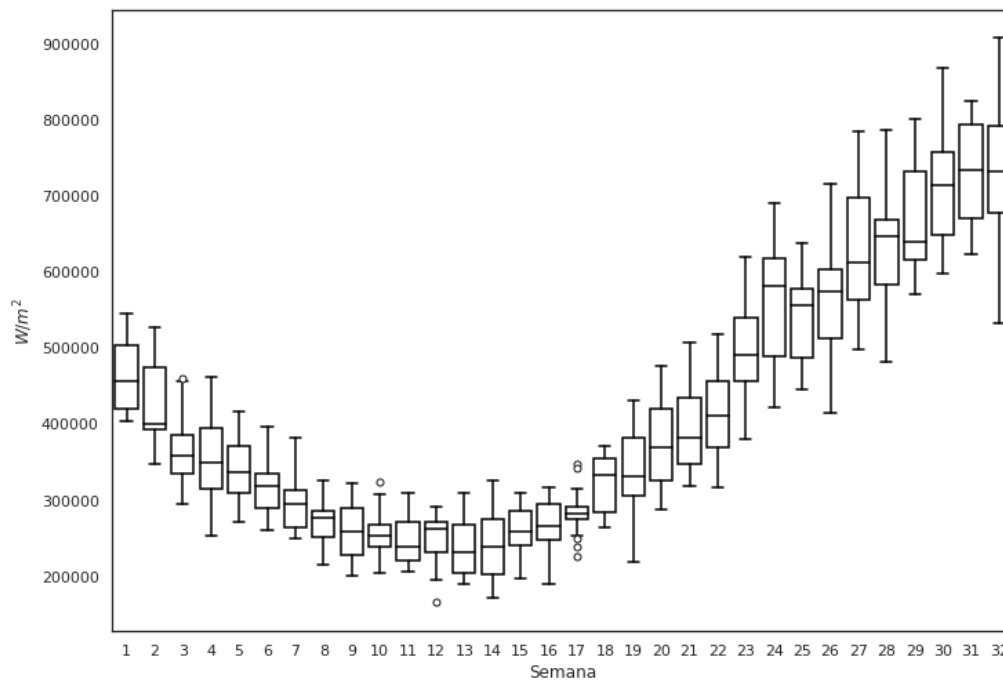


Figura 5.6: Diagrama de cajas para la radiación neta en superficie. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

NAO: Siendo un fenómeno atmosférico no anual podemos observar que presenta gran variabilidad durante todas las semanas.

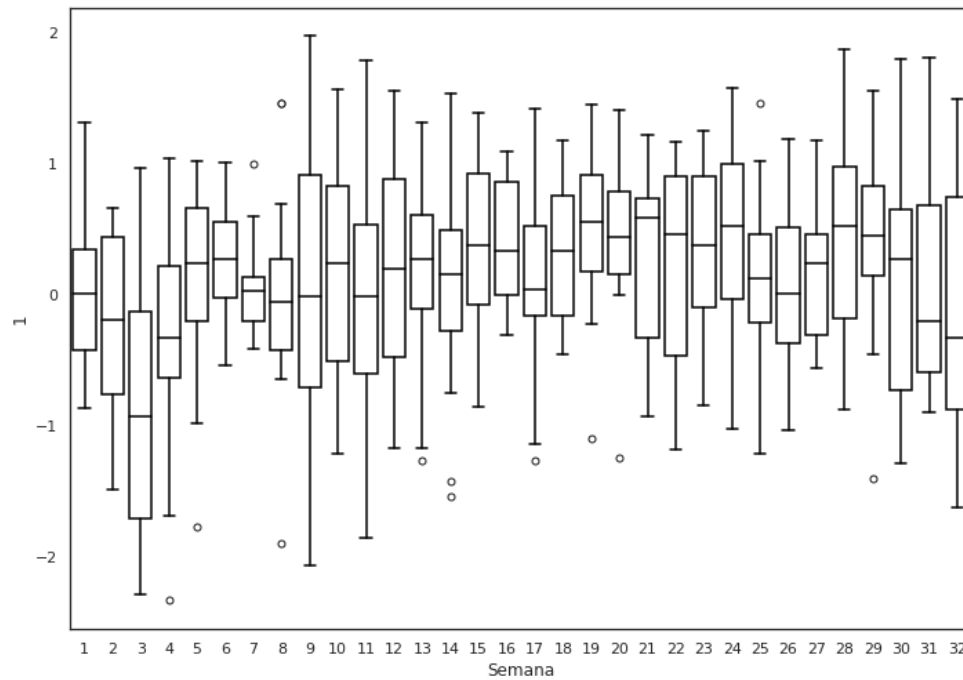


Figura 5.7: Diagrama de cajas para el NAO. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

Niño: Este evento climático no anual presenta mayor variabilidad a principios y mediados de temporada, donde dependiendo del año podremos observar las anomalías en el clima provocadas por dicho evento.

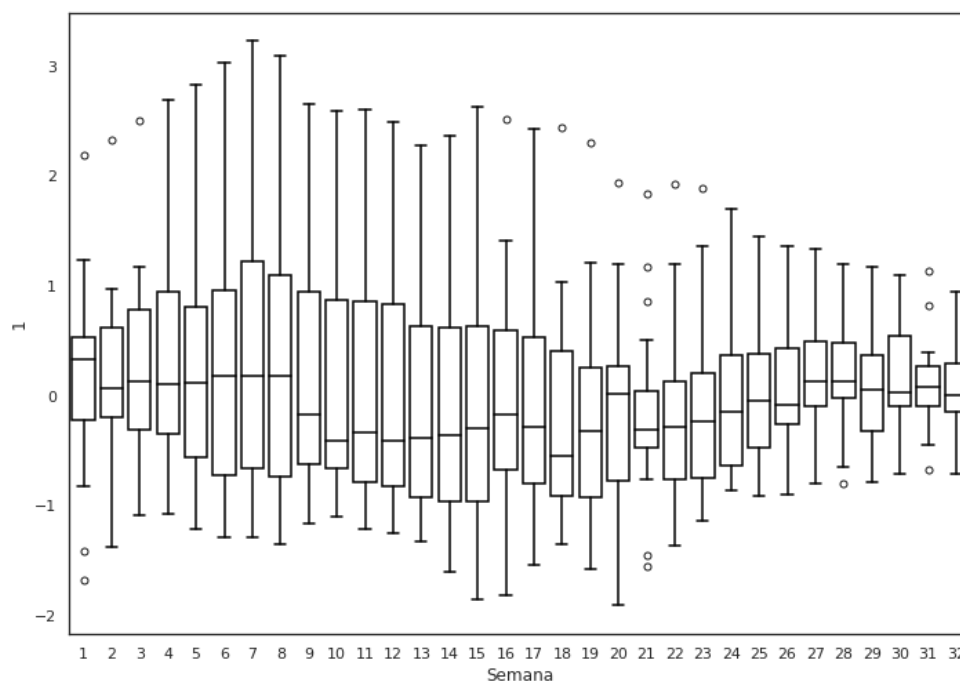


Figura 5.8: Diagrama de cajas para el fenómeno atmosférico *El Niño*. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

5.1.2. Variables de salida

Gripe: La estacionalidad es clara en los datos de la gripe. Podemos observar la presencia de muchos valores anómalos a principios de temporada, relacionados con los años en los que el pico de epidemia se produce antes, también podemos observar lo mismo a finales de la misma. La mayor variabilidad la encontramos a mitad de temporada donde el rango intercuartil es mayor, ya que no todos los años tenemos la misma tasa de infectados.

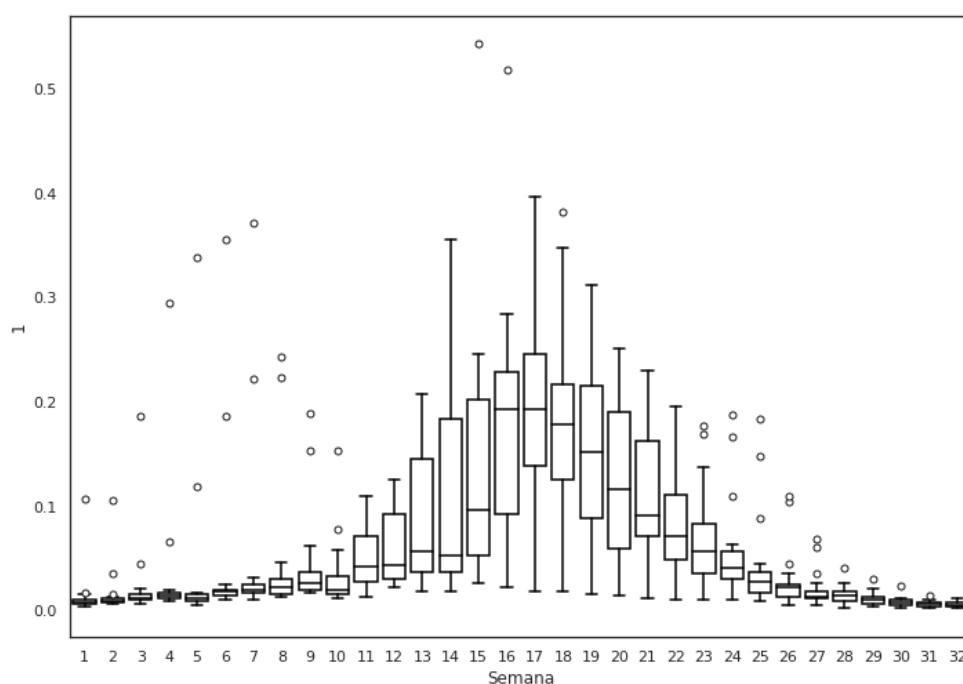


Figura 5.9: Diagrama de cajas para las temporadas de gripe. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

Beta: Los valores de beta calculados a partir de los datos de gripe¹, resultan más impredecibles a principios de temporada, donde vemos un rango intercuartil mayor y una mayor oscilación de la mediana, es posible que esta impredecibilidad produzca en determinadas temporadas una evolución anómala para la epidemia. A final de temporada podemos observar que los valores para la mediana se estabilizan en un rango de valores menores que 1². Esto es lógico que suceda ya que a final de temporada el número de contagios descende, y esto corresponde con valores menores que 1. Recordamos que el crecimiento de los infectados se ocasiona por valores mayores que 1.

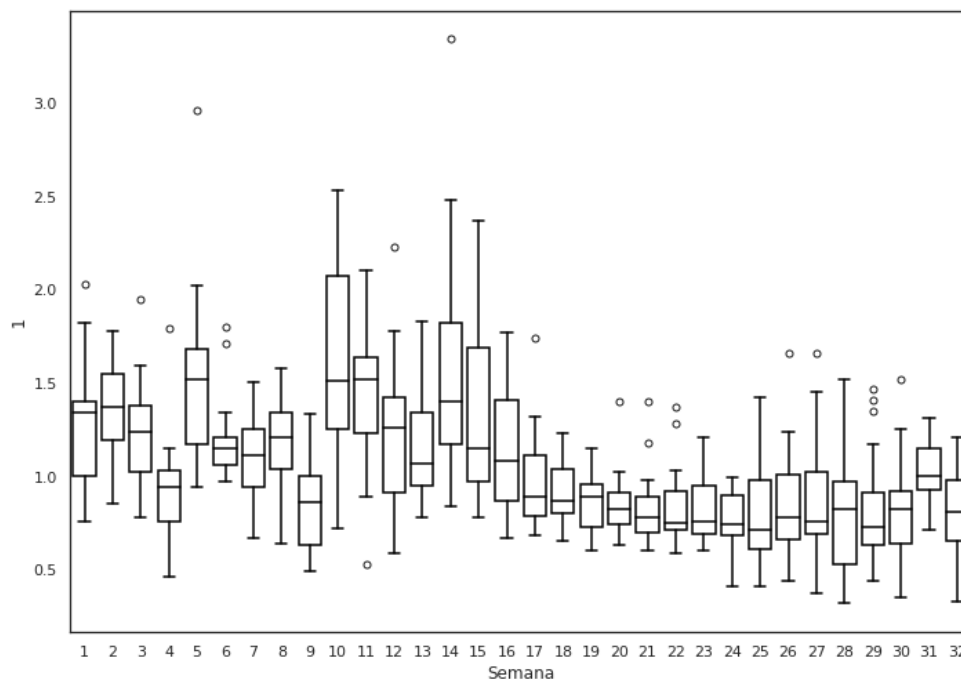


Figura 5.10: Diagrama de cajas para la variable *beta* del modelo SIR. La primera semana del segundo año de la temporada empieza a partir de la semana 14.

¹Véase la Sección 6.2 para los detalles del cálculo de dichos valores.

²Nótese que en los datos semanales $\gamma = 1$, por lo que $R(t) = \beta(t)$

5.2. Matriz de Correlaciones

Esta matriz nos indicará la intensidad con la que están correlacionadas las variables de nuestro conjunto de datos. Nos ayudará a corroborar algunas observaciones hechas en la Sección 5.1.

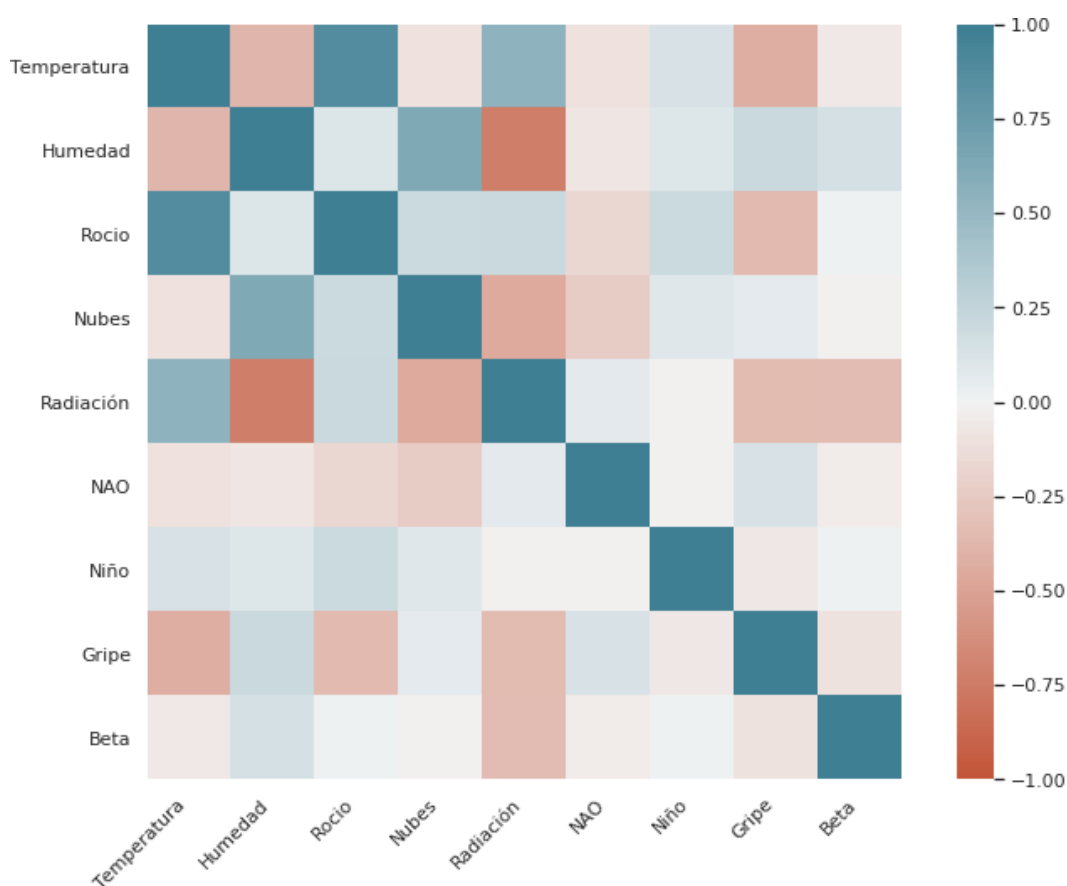


Figura 5.11: Matriz de correlaciones de las diferentes variables del conjunto de datos.

Podemos sacar diversas conclusiones a partir de dicha matriz:

- Tal y como era de esperar, existe una correlación fuerte y positiva entre la temperatura general y la temperatura de rocío.
- La radiación tiene una correlación fuerte y negativa con la humedad, y positiva con la temperatura.
- Podríamos considerar que el porcentaje de nubes tiene una correlación alta con respecto

a la humedad.

- Existe una correlación negativa y débil, entre la variable de infectados (gripe) y la temperatura.
 - No parece haber una correlación fuerte en ningún sentido con las demás variables.
-

6. Metodología del Modelo y Resultados

Todas las variables de nuestro conjunto de datos evolucionan en el tiempo y la salida de la red depende no solo de los datos actuales que recibe sino también de los datos que recibió en el pasado. Ante esta problemática es lógico utilizar una de las herramientas del ML: Recurrent Neural Network.

Trataremos en este apartado de construir una arquitectura que nos permita realizar los experimentos oportunos sin sobrepasar las capacidades de nuestros equipos informáticos.

En la primera fase, desarrollaremos una arquitectura que tratará de predecir la evolución de la epidemia en base a las variables climáticas que poseemos. La siguiente fase, requerirá realizar de manera previa la resolución del sistema de ecuaciones diferencial SIR, una vez resuelto obtendremos los valores del parámetro $\beta(t)$ que incluiremos en nuestro conjunto de datos. Cambiaremos la salida de la red para que en lugar de predecir los infectados por gripe haga predicciones sobre dicha variable que regula la expansión de la epidemia. Con estos valores de β , que evolucionarán en el tiempo y cuyas predicciones sobre dicha evolución serán realizadas por la red, obtendremos las diversas gráficas correspondientes a la evolución de la epidemia.

Para la creación de la red neuronal, así como para el resto de análisis sobre los datos, se ha utilizado como entorno de desarrollo el software *Jupyter Notebook* que nos permite realizar experimentos y visualizar los datos o resultados en diferentes módulos. Utilizando *Tensorflow* y la API *Keras* se han desarrollado las diferentes arquitecturas, también ha sido de gran utilidad trabajar con la librería *pandas* y *numpy* para manejar los datos y darles un formato adecuado para poder procesarlos con los modelos del ML.

Todas las arquitecturas usadas son entrenadas mediante diferentes *batch size*, el cual es un hiperparámetro que define el número de ejemplos a partir del conjunto de entrenamiento, por los cuales la red actualizará sus parámetros internos en cada ciclo (epoch). Son elegidos al azar y en los siguientes apartados referidos a la experimentación y resultados se especificará el tamaño elegido que ofrecía los mejores resultados. Se han utilizado ajustes que previenen el sobreentrenamiento, como una parada de emergencia en caso de que la red tarde demasiado en mejorar sus resultados.

6.1. Modelo predictivo de los infectados por gripe

6.1.1. Construcción del modelo

Los primeros modelos solo tenían en cuenta la temperatura general y de rocío, la radiación y el porcentaje de nubes. Se seleccionó las unidades GRU por su eficiencia y bajo coste de recursos en comparación con LSTM. No obstante estos primeros modelos, debido a la inexperience eran demasiado voluminosos innecesariamente para la tarea que se estaba realizando con los datos disponibles, tal y como se puede observar en la Figura 6.1.

Todos los modelos utilizan, de forma similar los parámetros que se describen a continuación, siendo la elección de los mismos una decisión en base a los resultados obtenidos.

Función de error y/o de salida: *error logarítmico cuadrático medio*. Ya que no queremos que errores demasiado grandes sean penalizados de sobre manera y el rango de valores de las variables es alto.

Optimizador: Tras varias pruebas con diversos optimizadores se escogió *Adam*, conocido en inglés como “*adaptive moment estimation*”, un método adaptativo de aprendizaje que fue el que mejores resultados ofreció en general.

Métrica: Se utiliza el último valor de la secuencia predicha por la red para evaluar la calidad del resultado comparándolo con los valores reales.

Model: "sequential"

Layer (type)	Output Shape	Param #
gru (GRU)	(None, None, 50)	8250
dropout (Dropout)	(None, None, 50)	0
gru_1 (GRU)	(None, None, 50)	15150
dropout_1 (Dropout)	(None, None, 50)	0
gru_2 (GRU)	(None, None, 50)	15150
dropout_2 (Dropout)	(None, None, 50)	0
gru_3 (GRU)	(None, None, 50)	15150
dropout_3 (Dropout)	(None, None, 50)	0
time_distributed (TimeDistri	(None, None, 1)	51

Total params: 53,751
 Trainable params: 53,751
 Non-trainable params: 0

Figura 6.1: Primera arquitectura diseñada para la predicción de la epidemia. Experimento 1.

Tras un aprendizaje más profundo, el volumen de capas de la arquitectura fue reducido y optimizado, consiguiendo resultados similares o mejores incluso. En la Figura 6.2 podemos ver los parámetros totales, reducidos considerablemente.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, None, 35)	595
gru (GRU)	(None, None, 35)	7455
time_distributed (TimeDistri	(None, None, 5)	180

Total params: 8,230
 Trainable params: 8,230
 Non-trainable params: 0

Figura 6.2: Estructura de las últimas arquitecturas diseñadas para la predicción de la epidemia. Experimento 2.

También se puede observar en la Figura 6.2 que se han incluido dos capas nuevas que optimizan el entrenamiento de la red:

Conv1D: Que ayudará a la red a aprender diferentes patrones cortos dentro de las secuencias.

TimeDistributed: Que nos permite devolver secuencias enteras en lugar de valores simples.

Otras arquitecturas trataban de combinar las primeras versiones y las últimas, tal y como se puede observar en la Figura 6.3. Se obtuvieron con ellas resultados óptimos.

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
gru (GRU)	(None, None, 120)	46080
dropout (Dropout)	(None, None, 120)	0
gru_1 (GRU)	(None, None, 120)	86760
dropout_1 (Dropout)	(None, None, 120)	0
time_distributed (TimeDistri	(None, None, 5)	605

```

Total params: 133,445
Trainable params: 133,445
Non-trainable params: 0

```

Figura 6.3: Combinación de las arquitecturas diseñadas para la predicción de la epidemia. Experimento 3.

6.1.2. Experimentos y resultados

En esta subsección se van a exponer los resultados de los diferentes experimentos que se han realizado con la red que predice directamente los infectados por el virus. Se realizan tres comparativas, véase:

Experimento 1: Corresponde a la primera etapa de experimentos en las que encontramos una arquitectura voluminosa por su número de unidades o bloques GRU.

Experimento 2: Corresponde a la etapa media de esta fase, se caracteriza por el uso de arquitecturas menos voluminosas y por el uso de capas convolucionales.

Experimento 3: Corresponde a los últimos experimentos en esta etapa, intentando perfeccionar los resultados anteriores combinando elementos de los anteriores experimentos como el aumento de unidades o el uso de capas convolucionales para el entrenamiento de la red.

Experimento	Parámetros totales entrenables	Tiempo de entrenamiento	Número de muestras por ciclo	Error de validación
1	53751	5min 10s	250	0.0035
2	8230	1min 22s	250	0.0034
3	133445	3min 50s	175	0.0050

Tabla 6.1: Tabla de resultados resumido para la primera fase del proyecto sin el uso de la variable beta.

En la Figura 6.4 podemos observar cómo los resultados de test para los tres experimentos relevantes correspondientes a esta etapa. Estos resultados se obtienen a partir del conjunto de test o validación, que es una temporada de gripe separada del resto que se han utilizado para entrenar la red. En particular, la de la temporada 2017/2018. Los parámetros de las diferentes redes se ajustan conforme al conjunto de entrenamiento¹, nunca al conjunto de test o validación.

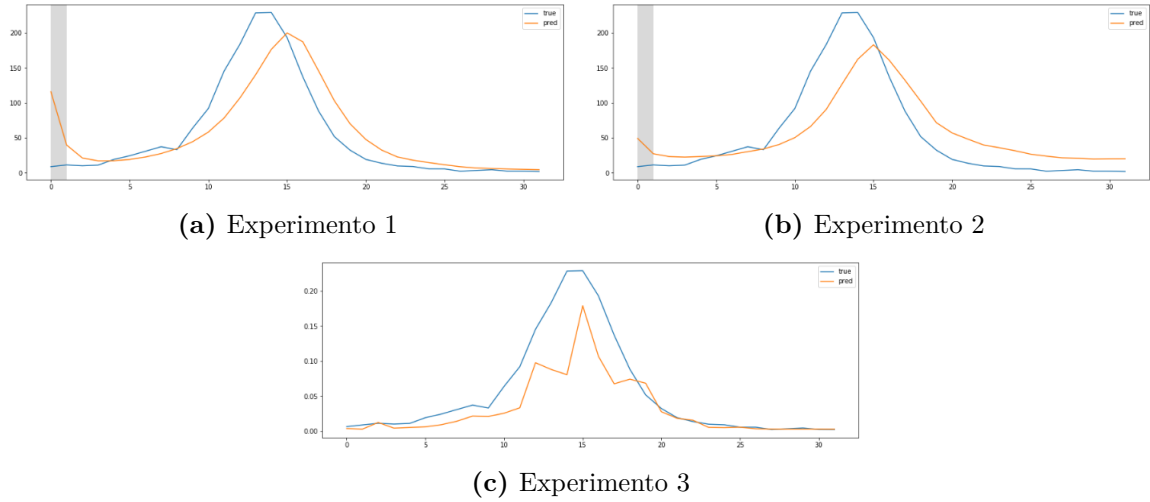


Figura 6.4: Resultados de test sin el uso de la variable beta. Leyenda: "true" para los datos reales, "pred" para los datos predichos por la red.

Se puede comprobar que podemos obtener resultados muy similares sin necesidad de crear una red demasiado voluminosa, ya que el experimento 2 contiene menos parámetros entrenables, no obstante la curva predicha para la epidemia en ambas redes (línea naranja) está desfasada con respecto a la referida a los datos reales (línea azul). En el experimento 3 o última fase de este apartado podemos ver como se consigue eliminar dicho desfase a costa de hacer que oscilen más los datos que predice la red.

¹Véase Anexo II: Gráficas de resultados, para consultar los resultados de la red para el conjunto de entrenamiento.

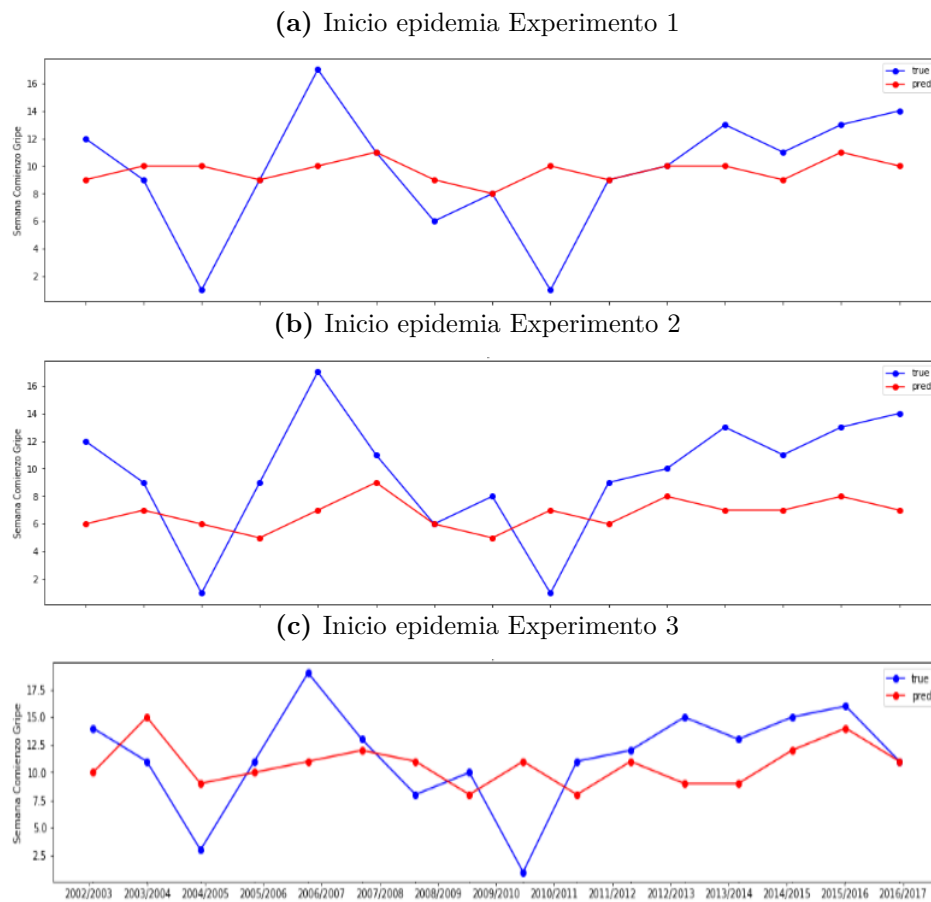


Figura 6.5: Inicios de la epidemia sin el uso de la variable beta.

Se considera el inicio de la epidemia cuando se tienen 40 infectados por cada 100000 habitantes. En la Figura 6.5 se puede observar el inicio de las epidemias que predice cada una de las redes con respecto al conjunto de entrenamiento. En todas ellas se aprecia una falta de variabilidad en los resultados de la red, observamos como el intervalo de valores, en el cual la epidemia tiene comienzo, se desplaza en sentido positivo o negativo según el experimento.

6.2. Modelo predictivo de la variable beta

6.2.1. Resolución del sistema de ecuaciones diferencial SIR

Para obtener los valores de β dentro de las ecuaciones diferenciales del SIR² tenemos que utilizar un procedimiento numérico, ya que no existe una solución analítica sencilla para este sistema de ecuaciones. Realizaremos una discretización del sistema de ecuaciones diferenciales del modelo SIR.

$$\frac{dR}{dt} = \gamma \cdot I \quad (2.2)$$

$$\frac{dS}{dt} = -\beta \cdot S \cdot I \quad (2.3)$$

$$\frac{dI}{dt} = \beta \cdot S \cdot I - \gamma \cdot I \quad (2.4)$$

mediante la aproximación de las derivadas por la fórmula de Euler,

$$f'(n) \approx \frac{(f_{n+1} - f_n)}{h},$$

Donde $f_n = f(n \cdot h)$, y $\{f_n\}_{n=0}^N$ es una discretización de la función f a intervalos de tiempo $h > 0$. De esta forma, discretizado el sistema se tiene:

$$\frac{S_{n+1} - S_n}{h} = -\beta \cdot S_n \cdot I_n \quad (6.1)$$

$$\frac{I_{n+1} - I_n}{h} = \beta \cdot S_n \cdot I_n - \gamma \cdot I_n \quad (6.2)$$

$$\frac{R_{n+1} - R_n}{h} = \gamma \cdot I_n \quad (6.3)$$

²Véase la Sección 2.2

Y despejando los términos $n + 1$:

$$S_{n+1} = S_n + h \cdot (-\beta \cdot S_n \cdot I_n) \quad (6.4)$$

$$I_{n+1} = I_n + h \cdot (\beta \cdot S_n \cdot I_n - \gamma \cdot I_n) \quad (6.5)$$

$$R_{n+1} = R_n + h \cdot (\gamma \cdot I_n) \quad (6.6)$$

- Siendo h la unidad de tiempo, en semanas tendremos $h = 1$.
- Durante el desarrollo se ha considerado β como un valor constante, no obstante si consideramos que varía en el tiempo, $\beta(t)$, solo hay que sustituir β por $\beta_n = \beta(n \cdot h)$, que es una discretización de $\beta(t)$.

6.2.2. Obtención de la variable Beta β y Gamma γ

La variable $\beta(n)$ se puede despejar de las ecuaciones 6.4 y 6.5 como:

$$\beta_n = \frac{-(S_{n+1} - S_n)}{S_n \cdot I_n} \quad (6.7)$$

$$\beta_n = \frac{I_{n+1} - I_n + \gamma \cdot I_n}{S_n \cdot I_n} \quad (6.8)$$

En la expresión 6.8 aparece el parámetro γ , dicho parámetro se determina como la inversa de la duración de la infección (véase Sección 2.2.1). La duración de la gripe según las observaciones dura de media una semana (Ghebrehewet et al., 2016), por tanto el valor de γ debe ser cercano a 1. Con el valor correcto de gamma, las dos estimaciones de β deberían ser iguales. No obstante, como dichas expresiones vienen de discretizar las ecuaciones del modelo SIR, no sería extraño contemplar diferencias entre ambas estimaciones. No obstante, es de esperar que las diferencias entre ambas expresiones se minimice con el valor correcto de γ . Por tanto, intentaremos calcular el valor de γ que minimice las diferencias entre los valores de β calculados a partir de las expresiones 6.7 y 6.8. La Figura 6.6 muestra el resultado de dicho cálculo. **El mínimo se alcanza para $\gamma = 0.96$.**

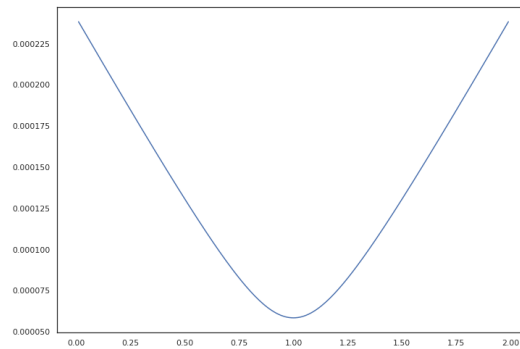


Figura 6.6: Buscando el valor de gamma adecuado.

6.2.3. Construcción del modelo

La construcción de los diferentes modelos partió de forma similar a lo expuesto en la Sección 6.1, donde los primeros modelos eran más voluminosos que los finales. El tamaño de los ejemplos que le pasaba a la red o *batch size* fue aumentado y con ello se mejoraron los resultados, fue relevante este hecho ya que la variable beta oscila entre rangos mayores y la red necesitaba más ejemplos para conseguir un entrenamiento óptimo.

Model: "beta-V4"

Layer (type)	Output Shape	Param #
gru (GRU)	(None, None, 36)	4428
gru_1 (GRU)	(None, None, 36)	7884
gru_2 (GRU)	(None, None, 36)	7884
gru_3 (GRU)	(None, None, 36)	7884
time_distributed (TimeDistri	(None, None, 36)	1332
dropout (Dropout)	(None, None, 36)	0
time_distributed_1 (TimeDist	(None, None, 15)	555
Total params: 29,967		
Trainable params: 29,967		
Non-trainable params: 0		

Figura 6.7: Primera arquitectura diseñada para la predicción de β . Experimento 4

Model: "beta-V4"

Layer (type)	Output Shape	Param #
gru (GRU)	(None, None, 35)	4200
gru_1 (GRU)	(None, None, 35)	7455
gru_2 (GRU)	(None, None, 35)	7455
gru_3 (GRU)	(None, None, 35)	7455
time_distributed (TimeDistri	(None, None, 35)	1260
dropout (Dropout)	(None, None, 35)	0
time_distributed_1 (TimeDist	(None, None, 5)	180

Total params: 28,005
 Trainable params: 28,005
 Non-trainable params: 0

Figura 6.8: Estructura de las últimas arquitecturas diseñadas para la predicción de β . Experimento 5.

Model: "beta-V6"

Layer (type)	Output Shape	Param #
gru (GRU)	(None, None, 120)	44640
gru_1 (GRU)	(None, None, 120)	86760
dropout (Dropout)	(None, None, 120)	0
time_distributed (TimeDistri	(None, None, 1)	121

Total params: 131,521
 Trainable params: 131,521
 Non-trainable params: 0

Figura 6.9: Combinación de las arquitecturas diseñadas para la predicción de β . Experimento 6.

6.2.4. Experimentos y resultados

Se van a exponer los resultados de los diferentes experimentos que se han realizado con la red que predice los valores de $\beta(t)$, que son usados posteriormente junto con las ecuaciones 6.4, 6.5 y 6.6 para realizar las predicciones sobre la evolución de la epidemia. Al igual que en la Sección 6.1.2, se han realizado experimentos en tres etapas donde se han probado diferentes configuraciones para la red. Cabe destacar que la experiencia ganada en los anteriores experimentos donde no se calculaba la variable $\beta(t)$ fue de gran ayuda para poder obtener mejores resultados en esta fase del proyecto.

Experimento 4: Los experimentos usando una red más voluminosa en cuando a número de capas para predecir la variable $\beta(t)$ fueron los mejores en este caso, arrojando mejores resultados y predicciones más precisas para el conjunto de validación.

Experimento 5: Los experimentos para esta red eran prometedores y mejoraban en el caso de aumentar el número de muestras por ciclo, aunque sus resultados finales no eran tan óptimos como los realizados en los primeros experimentos.

Experimento 6: Este tipo de red aunque llevaba menos capas, se aumento el número de parámetros entrenables y con ello aumento también el tiempo de entrenamiento para llegar a un resultado óptimo, no obstante seguía sin mejorar los resultados arrojados en los primeros experimentos.

Experimento	Parámetros totales entrenables	Tiempo de entrenamiento	Número de muestras por ciclo	Error de validación
4	29967	17min 39s	3500	0.0052
5	28005	13min 31s	3500	0.0056
6	131551	2h 31min 6s	4000	0.0078

Tabla 6.2: Tabla de resultados resumido para la segunda fase del proyecto con el uso de la variable beta.

Además de las predicciones de la red ("*pred*", línea naranja) y datos reales ("*true*", línea azul), podemos observar en la Figura 6.10 como se han recalculado también las epidemias a partir de los resultados de $\beta(t)$ obtenidos³ ("*true recalculate*", línea verde). Nótese la mejora del Experimento 4 en comparación a los Experimentos 5 y 6, y con respecto a los tres Experimentos de la Sección 6.1.

³Véase Apartado 6.2.2

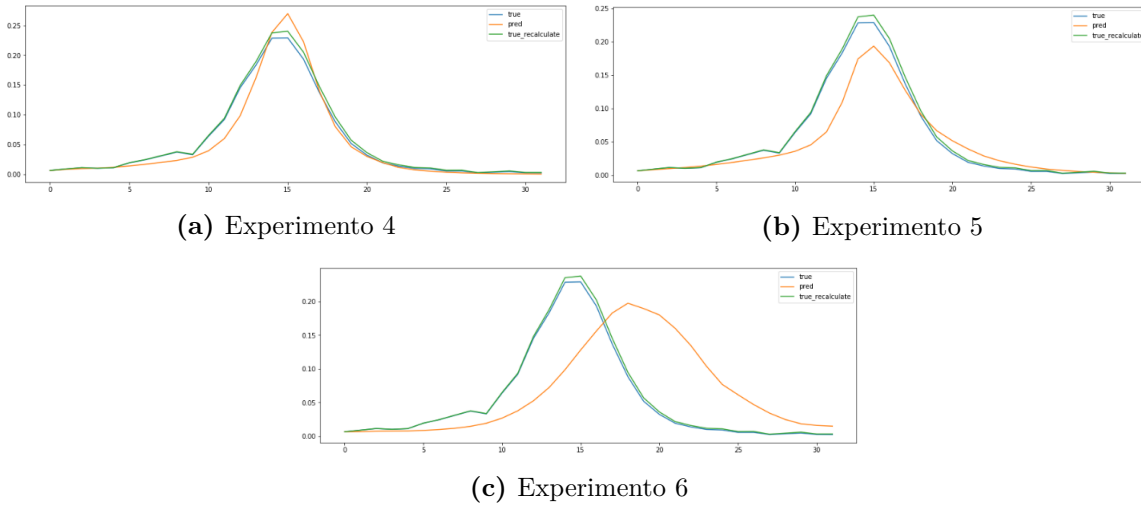


Figura 6.10: Resultados de test con el uso de la variable $\beta(t)$.

Experimento	Inicio (Semana)	Final (Semana)	Duración de la Epidemia
4	12	20	8 Semanas
5	12	21	9 Semanas
6	13	27	14 Semanas
Datos Reales	11	20	9 Semanas

Tabla 6.3: Tabla de Inicio, Final y Duración para los resultados del conjunto de validación o test con la variable β .

Para los experimentos realizados con respecto a la predicción de $\beta(t)$, además del inicio de las epidemias también se ha incluido la duración y el final de las mismas. Recordando que se considera el inicio de la misma cuando se superan los 40 infectados y el final cuando se baja de dicho valor. Podemos observar en la Figura 6.11 que para el conjunto de entrenamiento, la red ya obtiene mayor variabilidad en cuanto a la predicción del inicio de las epidemias. No ocurre lo mismo con el final de las epidemias, donde se ve claramente una menor variabilidad de la esperada en los resultados. Los experimentos 5 y 6 no contienen mejoras significativas con respecto al 4, tal y como podemos ver en la Figura 6.12 y 6.13. En el experimento 5 podemos considerar una menor exactitud en la predicción del final de las epidemias, por otro lado en el experimento 6 no se consigue realizar ninguna predicción con respecto al final de las epidemias durante toda la temporada. La red referida al experimento 4 sigue siendo la que mejores resultados ofrece.

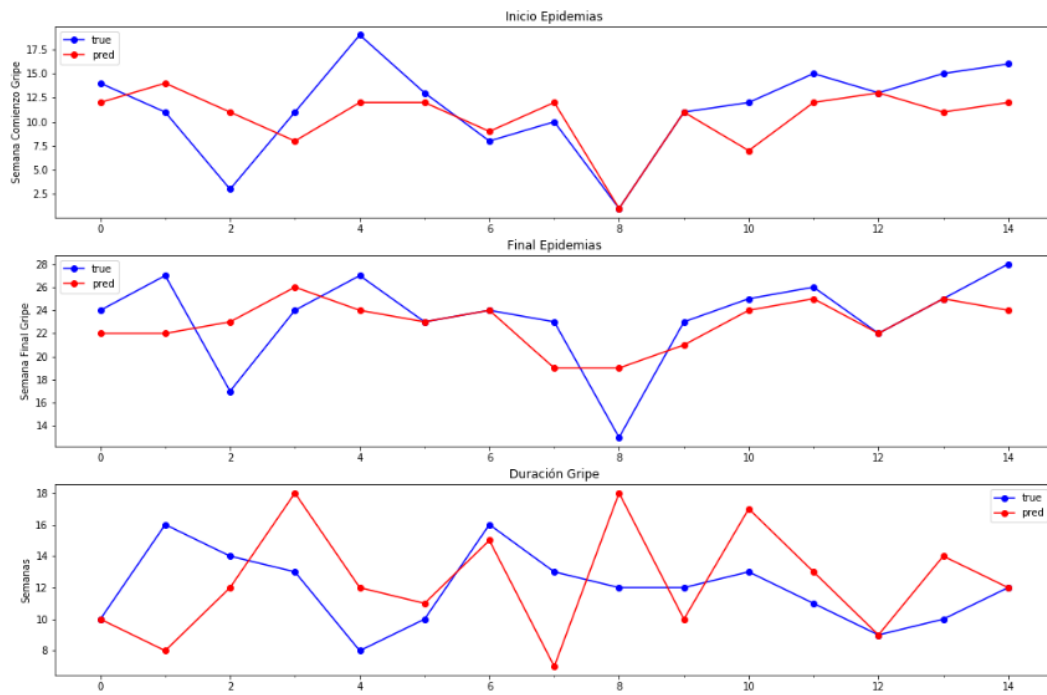


Figura 6.11: Inicio, final y duración para el Experimento 4

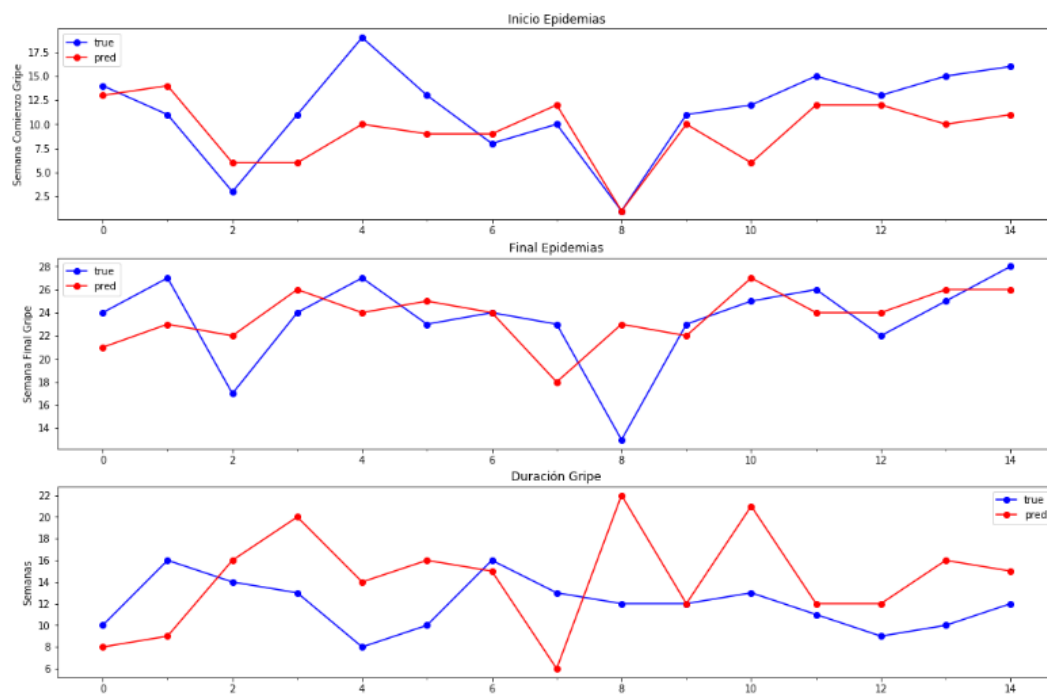


Figura 6.12: Inicio, final y duración para el Experimento 5

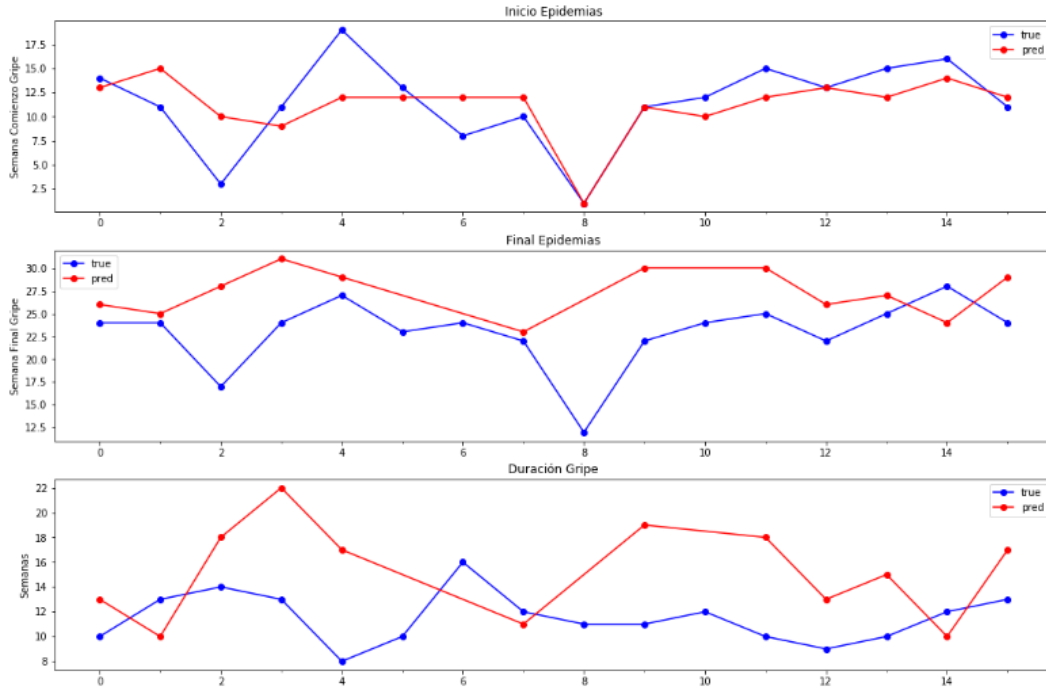


Figura 6.13: Inicio, final y duración para el Experimento 6

Generalmente la salida de nuestras redes obtiene buenos resultados para la variable $\beta(t)$, que como se puede ver en la Figura 6.14 sigue una tendencia similar a la dada por los datos reales dado el conjunto de validación.

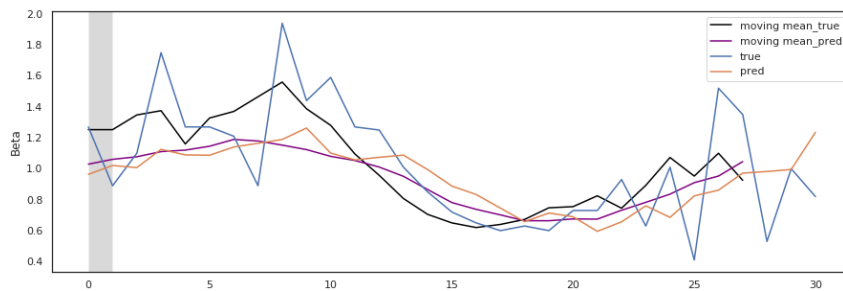


Figura 6.14: Variable $\beta(t)$. Conjunto de validación. Leyenda: "true" datos reales, "pred" predicciones de la red, "moving mean_true" media móvil de los datos reales y "moving mean_pred" media móvil de las predicciones de la red.

En la Sección B.2, dentro del Anexo II: Gráficas de resultados, podemos observar con más detalle en las diferentes figuras los resultados para el conjunto de entrenamiento.

7. Conclusiones

En este trabajo se ha podido comprobar experimentalmente la dependencia que tiene la evolución de la epidemia de gripe con respecto a las variables climáticas. El estudio del mismo ha tenido sus limitaciones debido a que la propagación del virus no solo depende de dichas condiciones climáticas, sino también de la capacidad de propagación y resistencia del virus propio de cada temporada, nivel de vacunación de cada temporada y efectividad de la misma.

Se han entrenado dos redes neuronales. La primera, enfocada en la predicción de la epidemia en sí; y la segunda, haciendo predicciones sobre el parámetro $\beta(t)$, el cual está relacionado con la tasa efectiva de reproducibilidad y determina la dinámica de la epidemia, con dicho parámetro se han calculado los infectados mediante el modelo SIR. En general, se han obtenido mejores resultados mediante la determinación del parámetro $\beta(t)$. Las razones son desconocidas en este punto de la investigación, se necesitan más experimentos y un estudio más profundo para esclarecer esta cuestión. A pesar de las limitaciones se han conseguido resultados prometedores haciendo predicciones de las epidemias estacionales estándar. Para temporadas con un inicio más o menos temprano de lo habitual también se han conseguido buenos resultados en algunos entrenamientos mediante el cálculo del parámetro $\beta(t)$ por parte de la red. El siguiente paso sería conseguir un híbrido de las redes con los resultados más óptimos para tener predicciones precisas tanto en las temporadas estándar como anómalas.

Los resultados prometedores nos hacen reflexionar sobre la posibilidad de profundizar más en este tema. Siendo necesario un estudio más detallado y pormenorizado, eliminando algunos datos climáticos y observando los resultados obtenidos, los cuales nos permitirían ver la influencia relativa de cada una de las variables climáticas¹. Otra posible mejora sería am-

¹En la Sección 8 se detallan algunas posibles mejoras del estudio.

pliar el estudio a otros países europeos e intentar encontrar similitudes y diferencias que nos permitan entender mejor el problema. Recordemos que a día de hoy no hay ningún tipo de información predictiva ni del inicio ni de la duración de las epidemias anuales de gripe. Cualquier información en este respecto sería de gran ayuda para planificar los recursos humanos y materiales del sistema nacional de salud.

El objetivo del proyecto se ha cumplido en el sentido de que ahora entendemos un poco mejor el problema y sabemos que podríamos mejorar los resultados, abriendo una posible línea de investigación. No obstante, debido a la complejidad del problema, el trabajo necesario para seguir avanzando sería más propio de una tesis doctoral que de un trabajo fin de grado. El objetivo de dicha tesis sería, por un lado, comprender mejor la dinámica de las epidemia estacionales, y por otro, diseñar un sistema de detección temprana de inicio y duración de brotes de gripe. El sistema se alimentaría de datos a tiempo real del modelo ERA5 y realizaría predicciones cada semana en función de las variables atmosféricas previas.

8. Proyectos futuros

En este apartado consideramos la posibilidad de ampliar este proyecto señalando experimentos que se podrían llevar a cabo.

Conjunto de datos :

- Aumentar los datos de la gripe interpolando los mismos a días. No haría falta realizar lo mismo con los datos climáticos, ya que los valores reales están almacenados de forma diaria.
- Añadir variables climáticas nuevas como el viento, el cual puede que afecte a la propagación de aerosoles.
- Comprobar la influencia relativa de cada variable ambiental, esto es, repetir los experimentos dejando fuera cada vez una variable ambiental. Comparar los errores con los experimentos en los que se usaban todas las variables. ¿Se mantiene dicha influencia en el tiempo?

Red Neuronal :

- Realizar estrategias de validación como *K-Fold Cross-Validation*.
- Modificaciones: realizarlas para infectados y para $\beta(t)$.
- Medición del error: medir la bondad del ajuste en cada caso respecto a infectados totales, mínimos cuadrados, inicio, fin, duración.

Bibliografía

- Anice C. Lowen, J. S. (2014). Roles of humidity and temperature in shaping influenza seasonality. *Journal of Virology*, 88(14).
- Blickle, K. (2020). Pandemics change cities: Municipal spending and voter extremism in germany, 1918-1933. *Federal Reserve Bank of New York Staff Reports*.(1).
- Cannel, J. J., y cols. (2006). Epidemic influenza and vitamin d. *Epidemiology and Infection*, 134, 1129–1140.
- Cannel, J. J., y cols. (2008). On the epidemiology of influenza. *Virology Journal*, 5, pp. doi: 10.1186/1743-422X-5-29.
- Castro, A., y cols. (2011). May the nao index be used to forecast rain in spain? *Atmósfera*, 24(3), 251–265.
- Cook, N. D. (1998). *Born to die: Disease and new world conquest, 1492–1650*. Cambridge: Cambridge University Press.
- D. Dalziel, B., Kissler, S., R. Gog, J., Viboud, C., N. Bjørnstad, O., y E. Metcalf, y. T. G. B., C. J. (2018). Urbanization and humidity shape the intensity of influenza epidemics in u.s. cities. *Science*, 362(6410), 75–79.
- Fuhrmann, C. (2010). The effects of weather and climate on the seasonality of influenza: What we know and what we need to know. *Geography Compass*, 4(7), 718–730.
- Ghebrehewet, S., MacPherson, P., y A., H. (2016). Influenza. *BMJ*. 2016; 355: i6258., Published online 2016 Dec 7. doi: 10.1136/bmj.i6258.

- Grima, C., y F.Borja, E. (2017). *Las matemáticas vigilan tu salud. modelos sobre epidemias y vacunas*. Next Door Publishers.
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras and tensorflow*. O'Reilly Media.
- Nunn Nathan, Q. N. (2010). The columbian exchange: A history of disease, food, and ideas. *Journal of Economic Perspectives*, 24(2), 163–188.
- Pollicott, M., Wang, H., y Weiss, H. H. (2012). Extracting the time-dependent transmission rate from infection data via solution of an inverse ode problem. *Journal of Biological Dynamics.*, 6(2), 509–523.
- Potter, C. (2001). A history of influenza. *Journal of Applied Microbiology.*, 91, 572–579.
- Pérez-Rubio Alberto, E. B. J., Platero Leticia. (2019). Gripe estacional en españa: carga clínica y económica y programas de vacunación. *Medicina Clínica*, 153(1), 16–17.
- Ropelewski, C. F., y Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the el niño/southern oscillation. *American Meteorological Society*, 115(8), 1606–1626.
- Solé, R., y Santiago F, E. (2019). *Viruses as complex adaptive systems*. Princeton University Press.
- T., T., S.Marinova, R., Omojola, J., y Jackson, M. (2014). Inverse problem for coefficient identification in sir epidemic models. *Computers Mathematics with Applications.*, 67(12), 2218–2227.
- von Neumann, J. (1966). *The theory of self-reproducing automata*. University of Illinois Press.
-

Lista de Acrónimos y Abreviaturas

API	Application Programming Interface.
CNN	Convolutional Neural Network.
DL	Deep Learning.
DNN	Deep Neural Network.
ECMWF	European Centre for Medium-Range Weather Fore- cast.
ERA5	European Research Area.
FEN	Fenómeno El Niño.
GRU	Gated Recurrent Unit.
KNMI	Koninklijk Nederlands Meteorologisch Instituut.
LSTM	Long Short-Term Memory.
ML	Machine Learning.
NAO	North Atlantic Oscillation.
NBR	Número Básico de Reproducción.
RENAVE	Red Nacional de Vigilancia Epidemiológica.
RNN	Recurrent Neural Network.
SIR	Susceptible Infectado Recuperado.
SOI	Southern Oscillation Index.
SVGE	Sistema de Vigilancia de la Gripe en España.

A. Anexo I: Gráficas de visualización de las variables del conjunto de datos

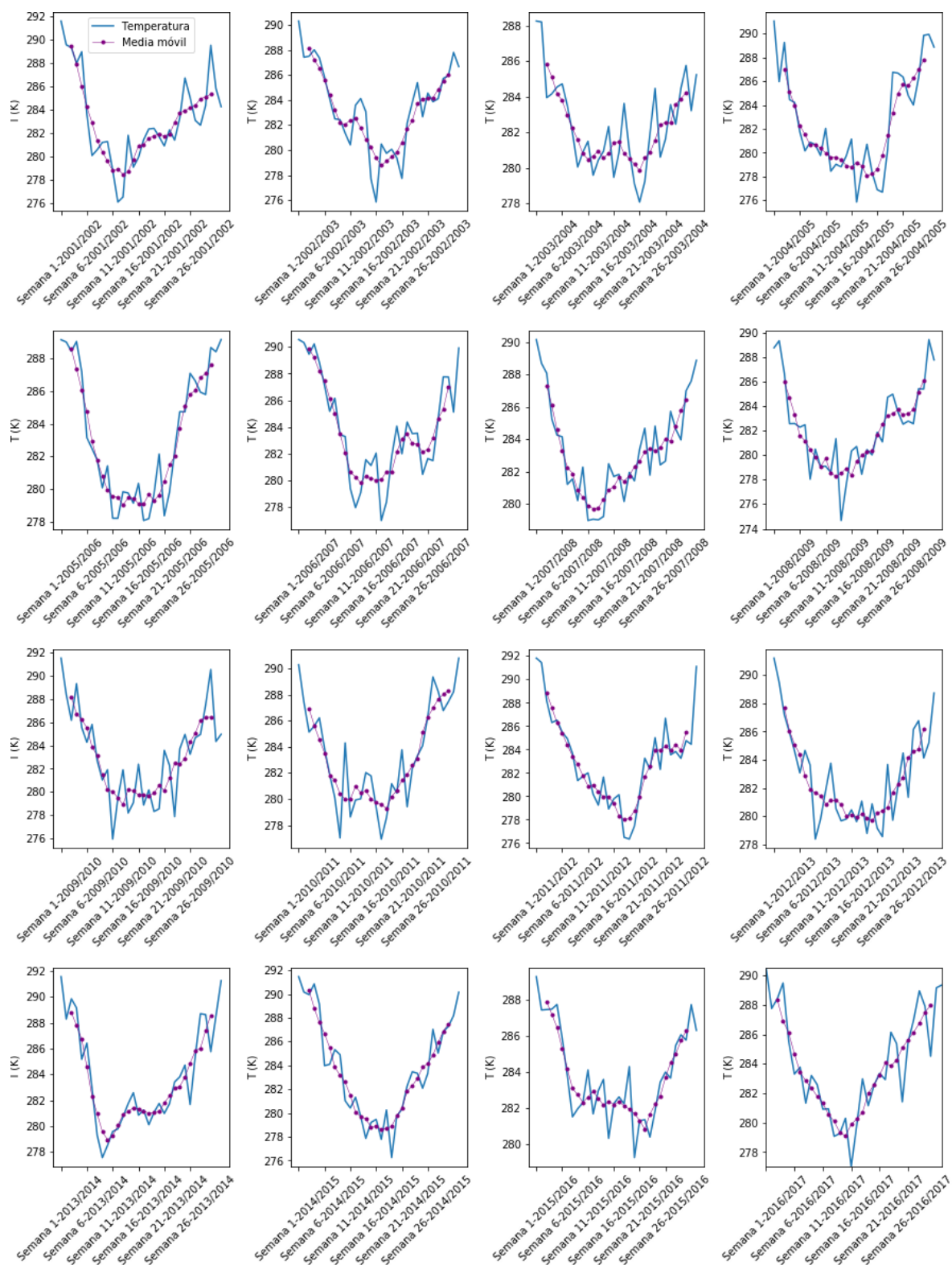


Figura A.1: Temperatura asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

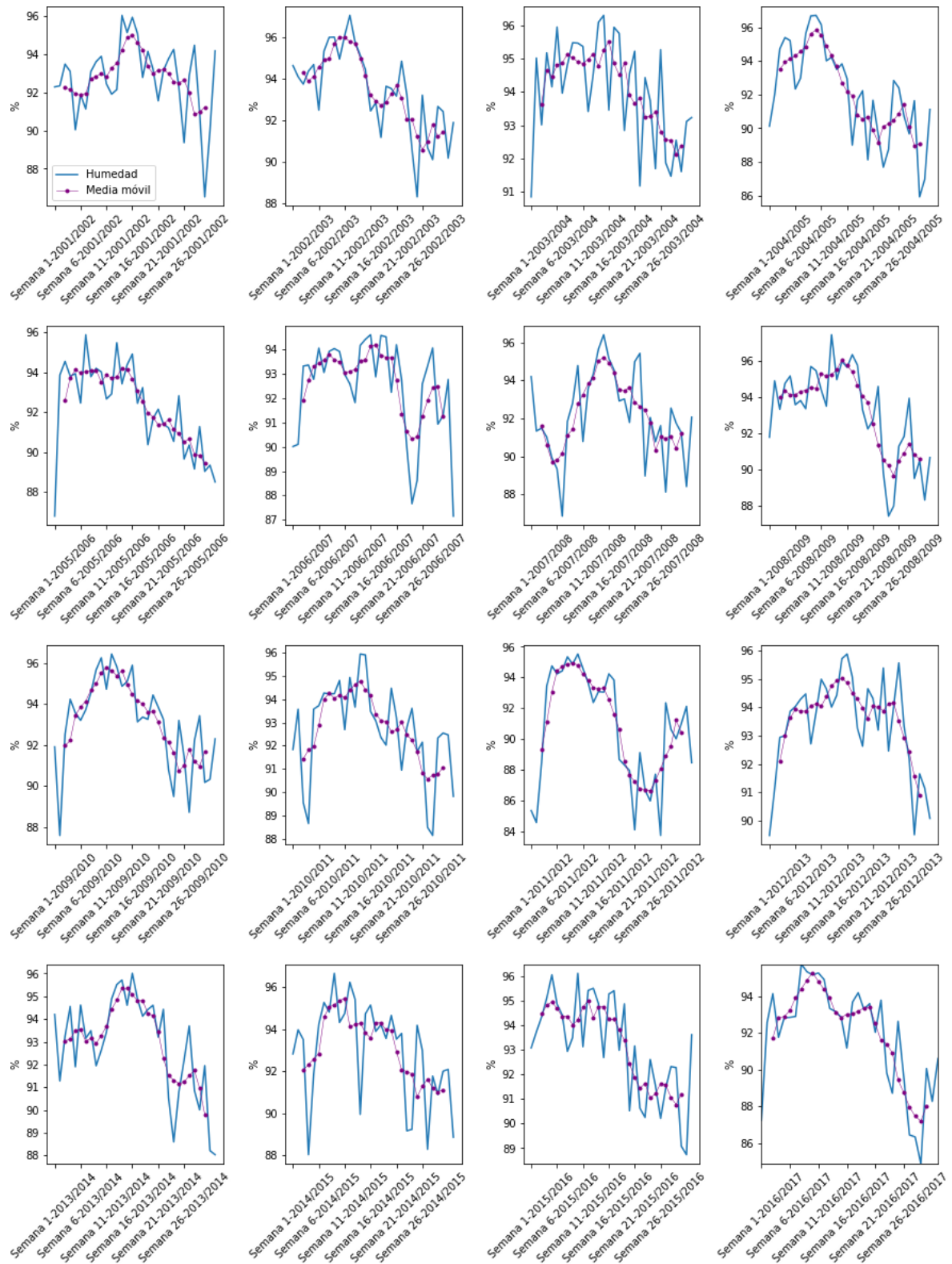


Figura A.2: Humedad asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

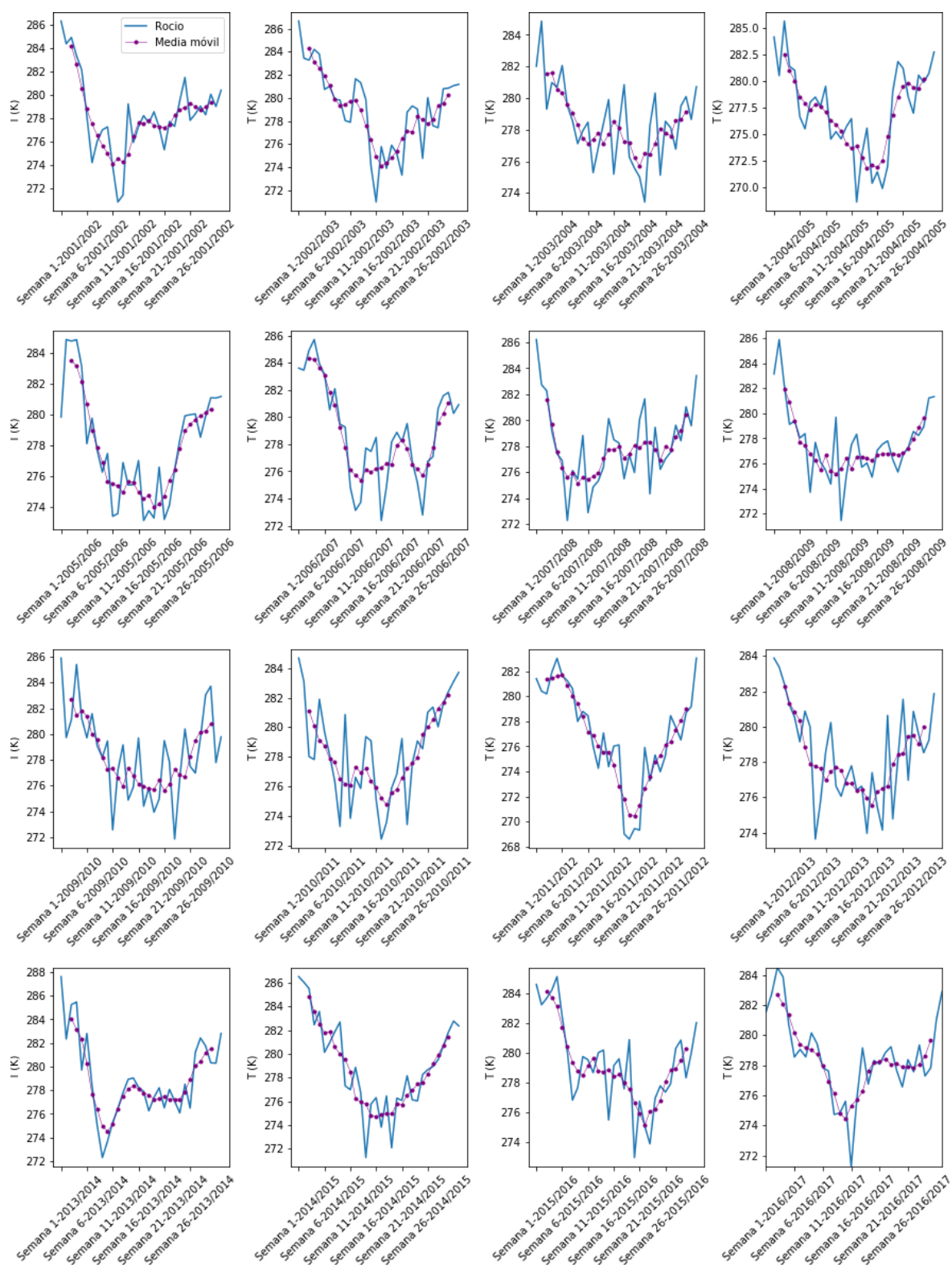


Figura A.3: Punto de rocío asociado a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

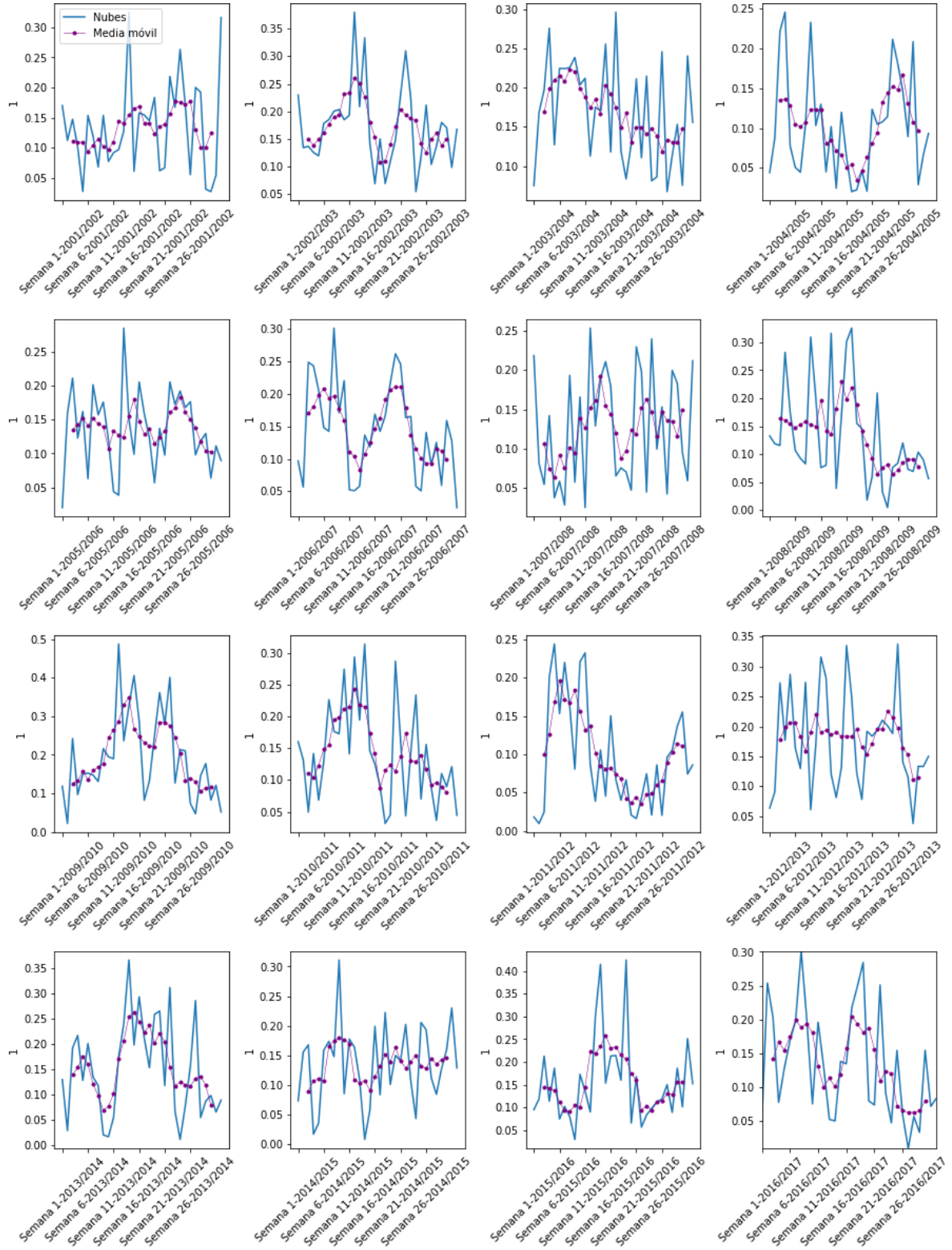


Figura A.4: Porcentaje de nubes asociado a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

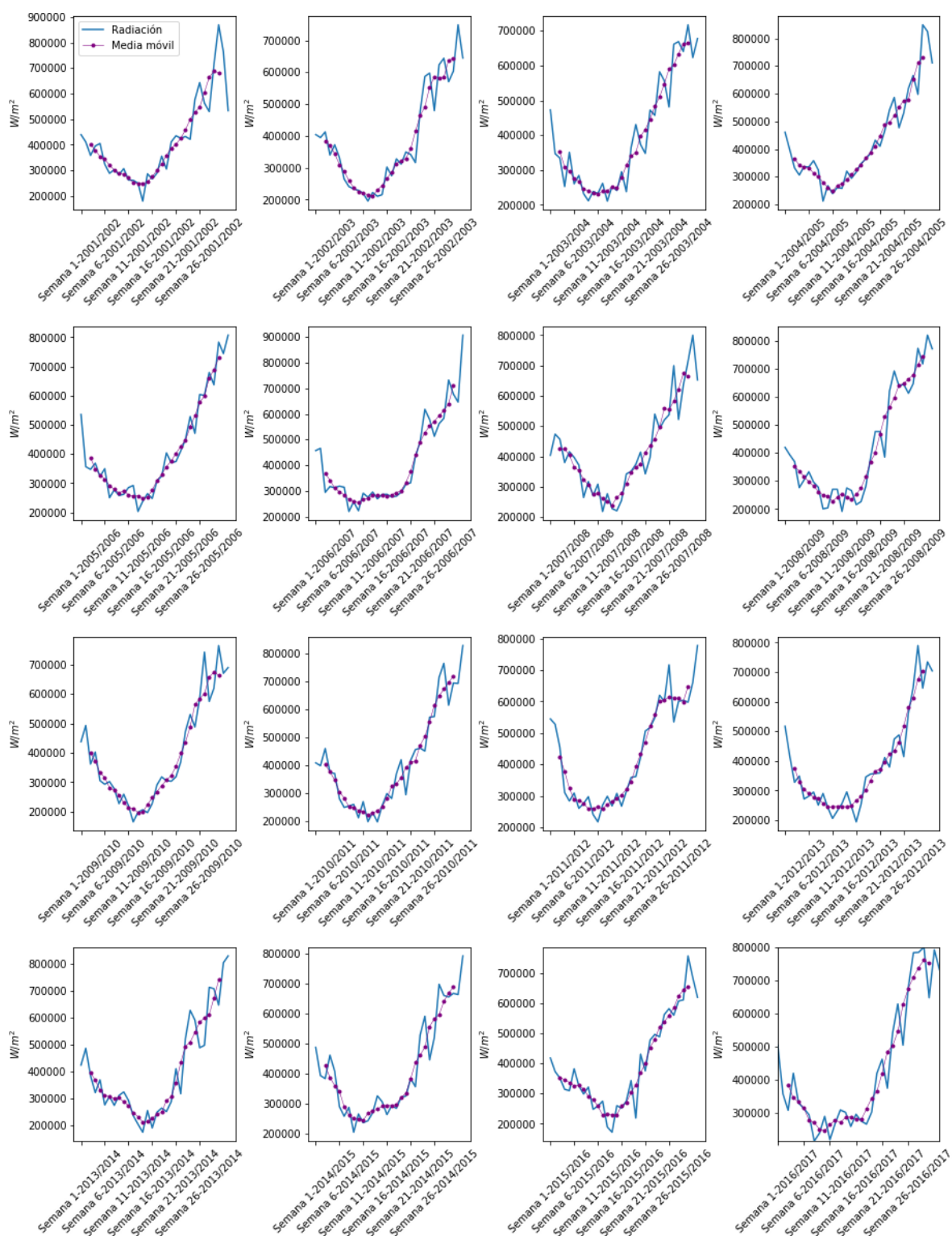


Figura A.5: Radiación neta en superficie asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

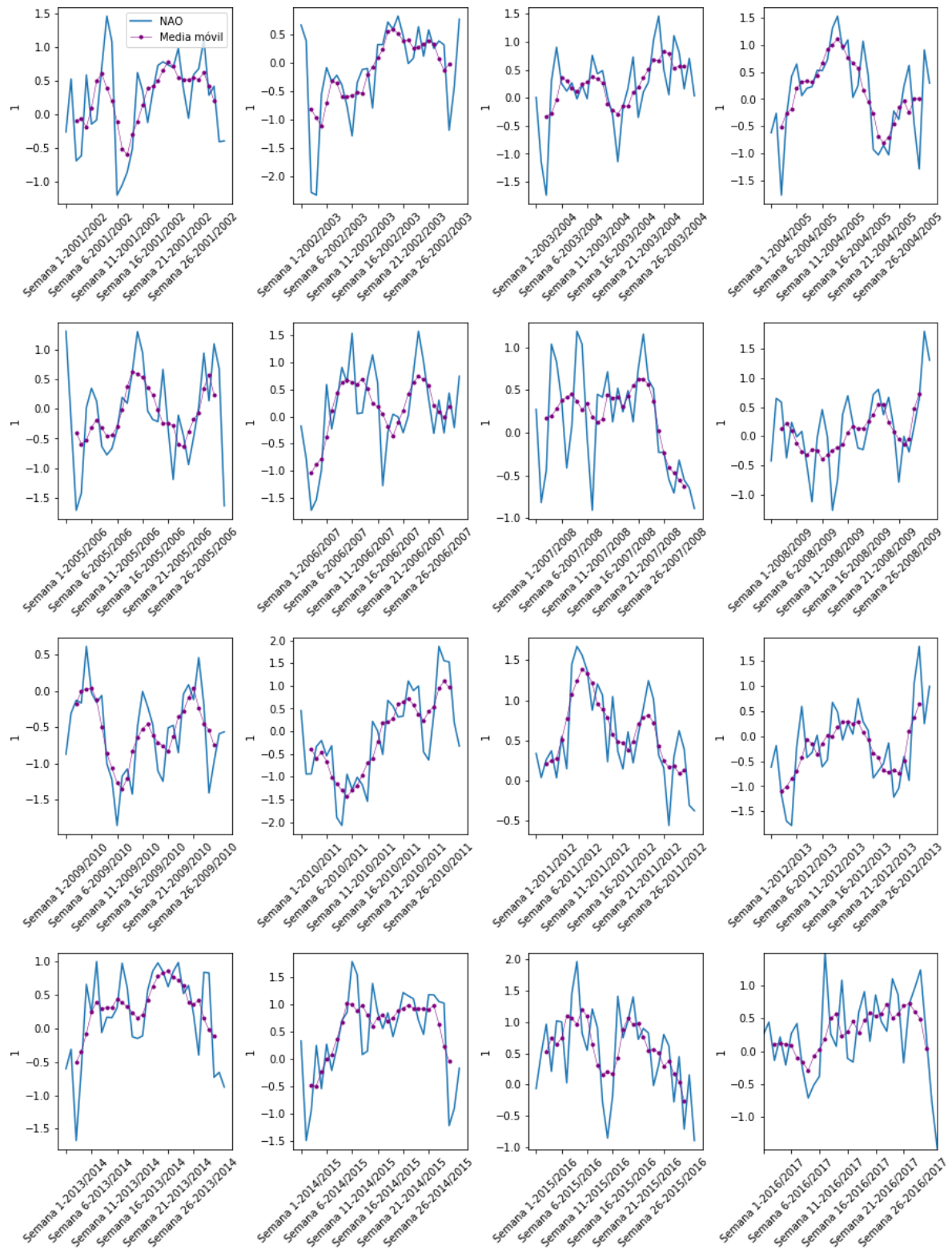


Figura A.6: NAO asociada a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

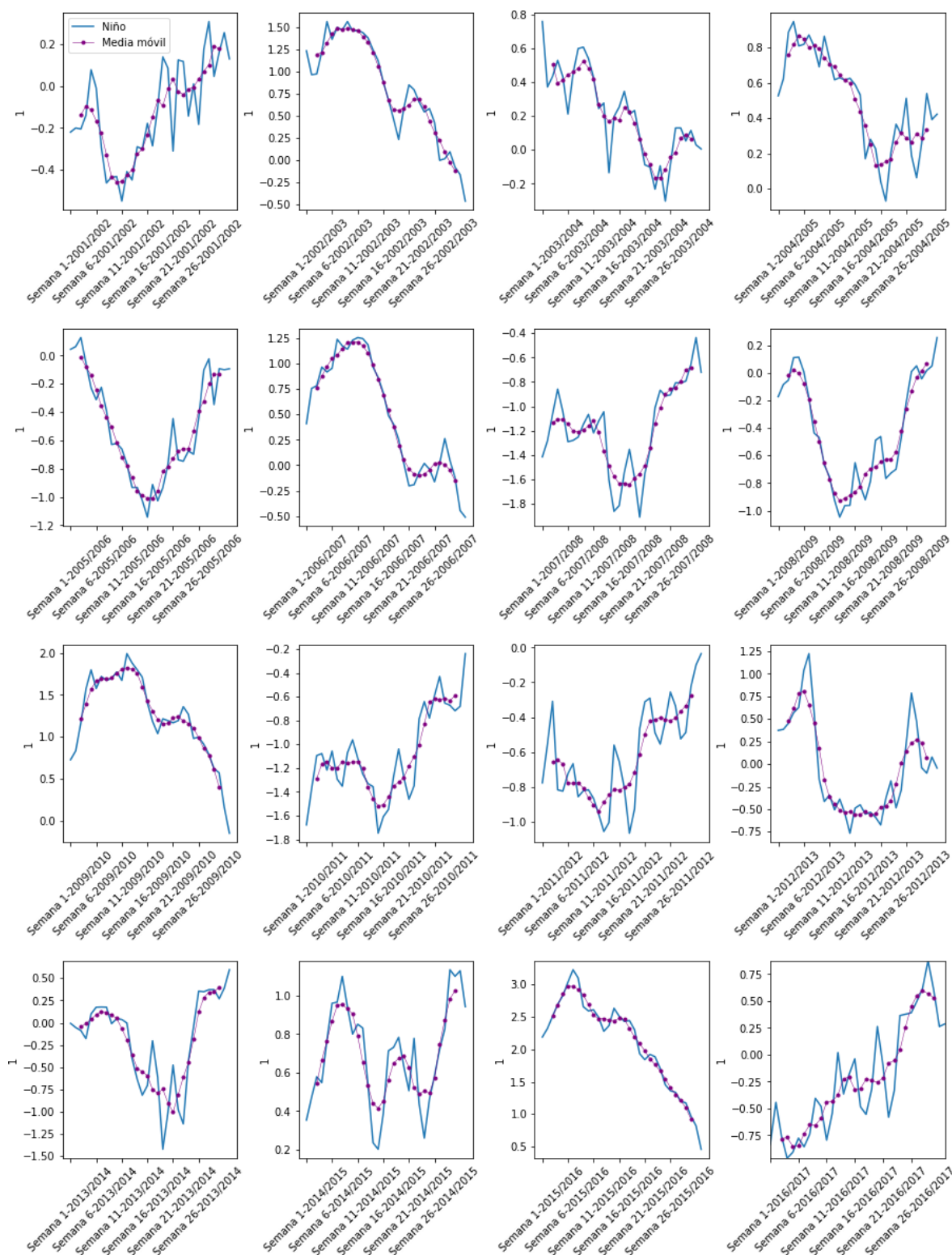


Figura A.7: El fenómeno de El Niño asociado a las diferentes temporadas. Media móvil ponderada a partir de 5 semanas.

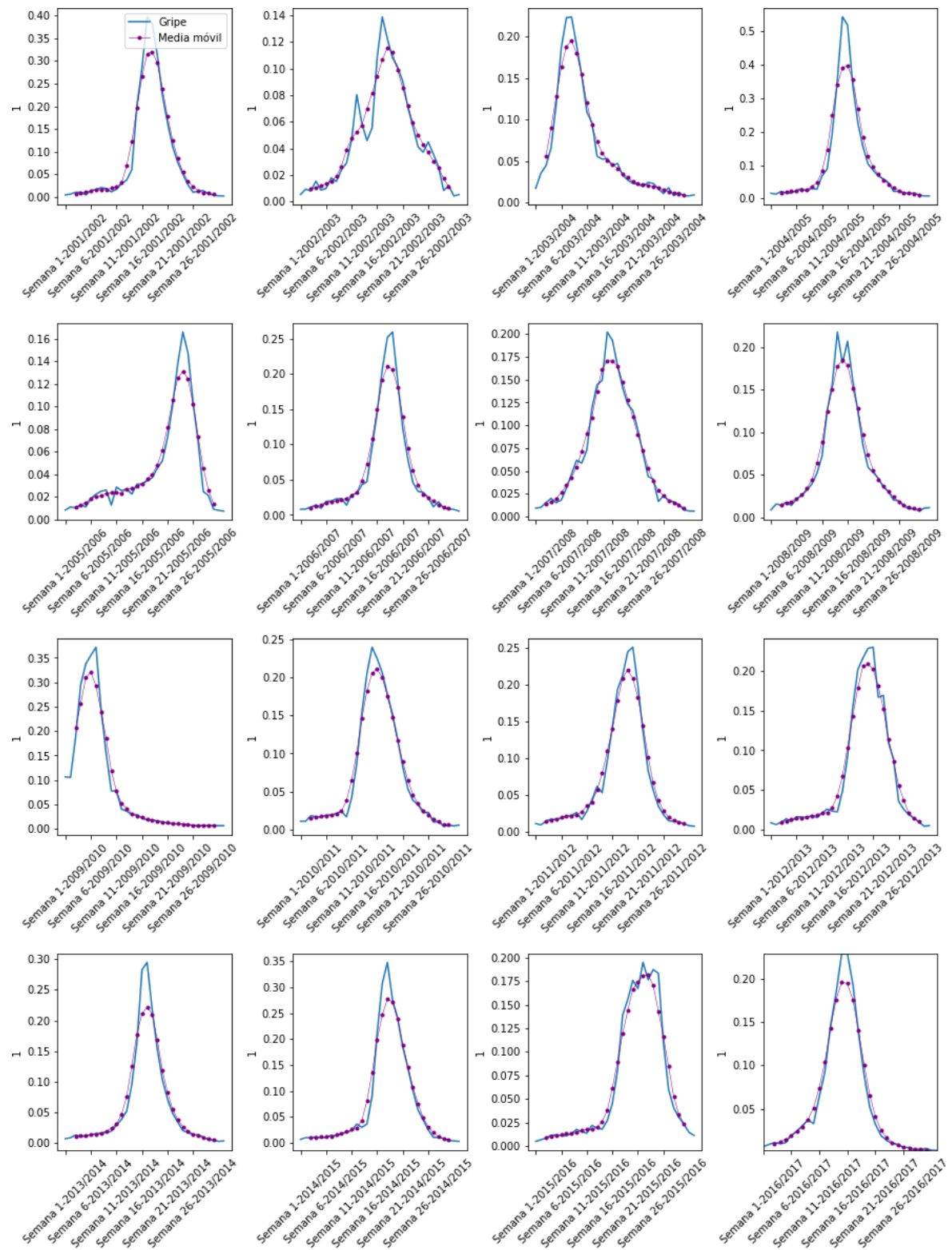


Figura A.8: Temporadas de Gripe. Media móvil ponderada a partir de 5 semanas.

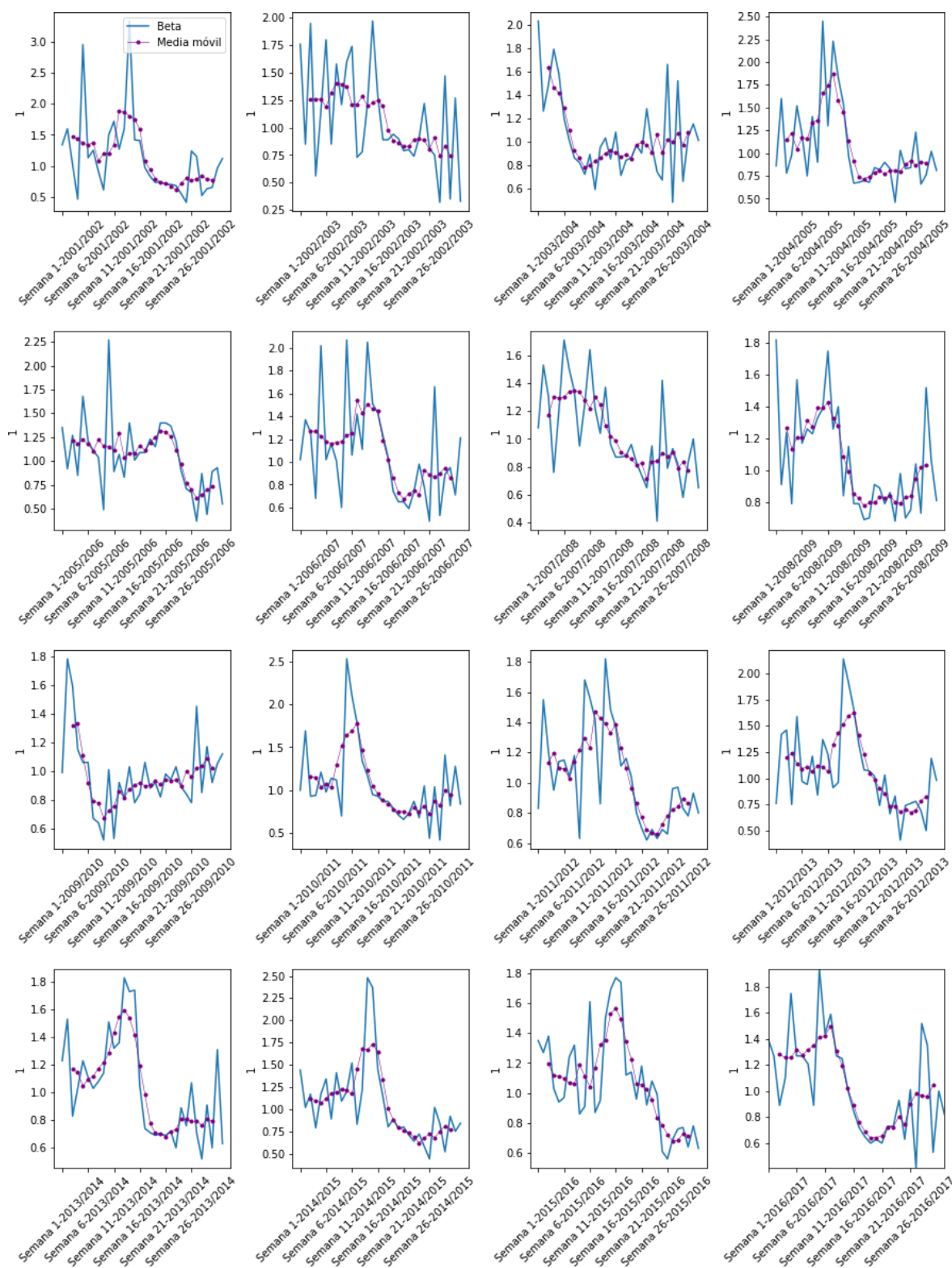


Figura A.9: La variable beta del modelo SIR, objetivo principal del proyecto. Media móvil ponderada a partir de 5 semanas.

B. Anexo II: Gráficas de resultados

B.1. Resultados modelo predictivo de infectados por gripe

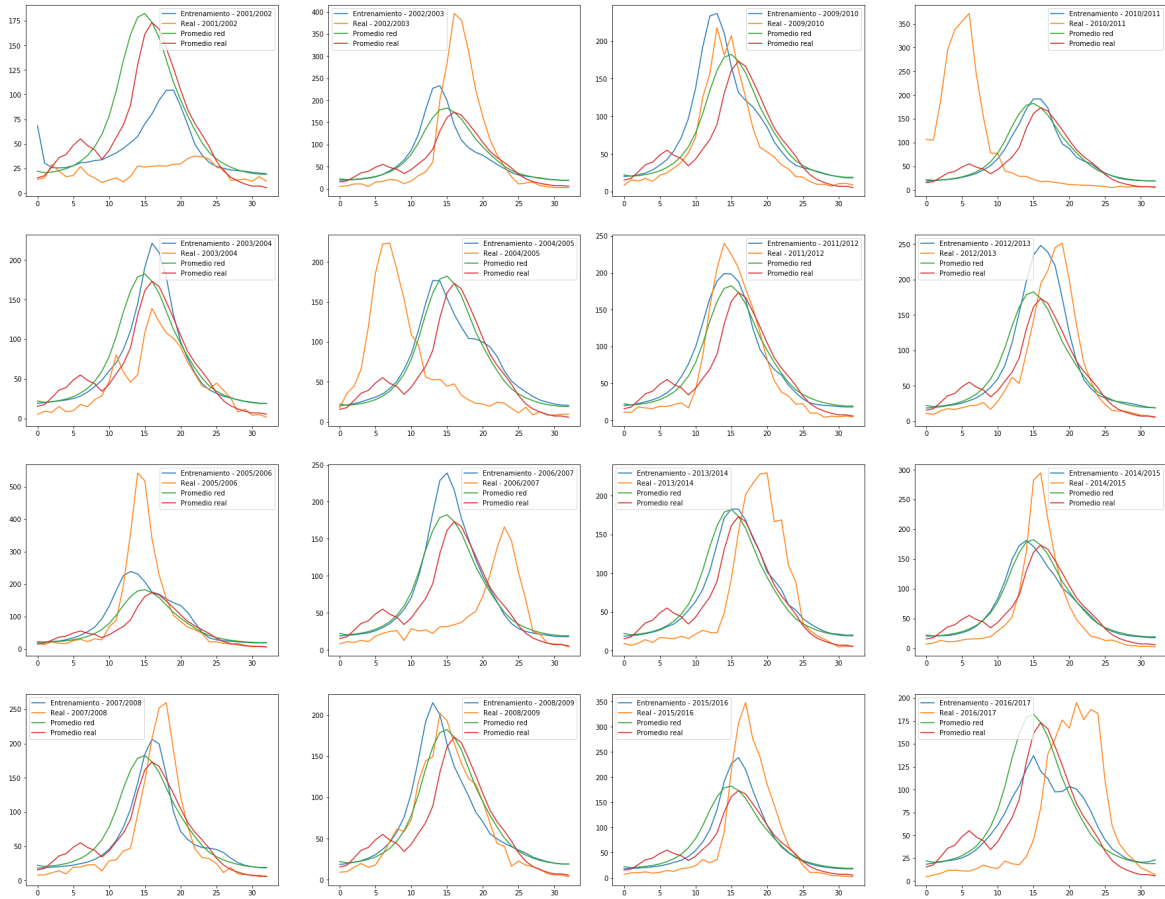


Figura B.1: Resultados de conjunto de entrenamiento. Modelo sin cálculo de $\beta(t)$. Experimento 1

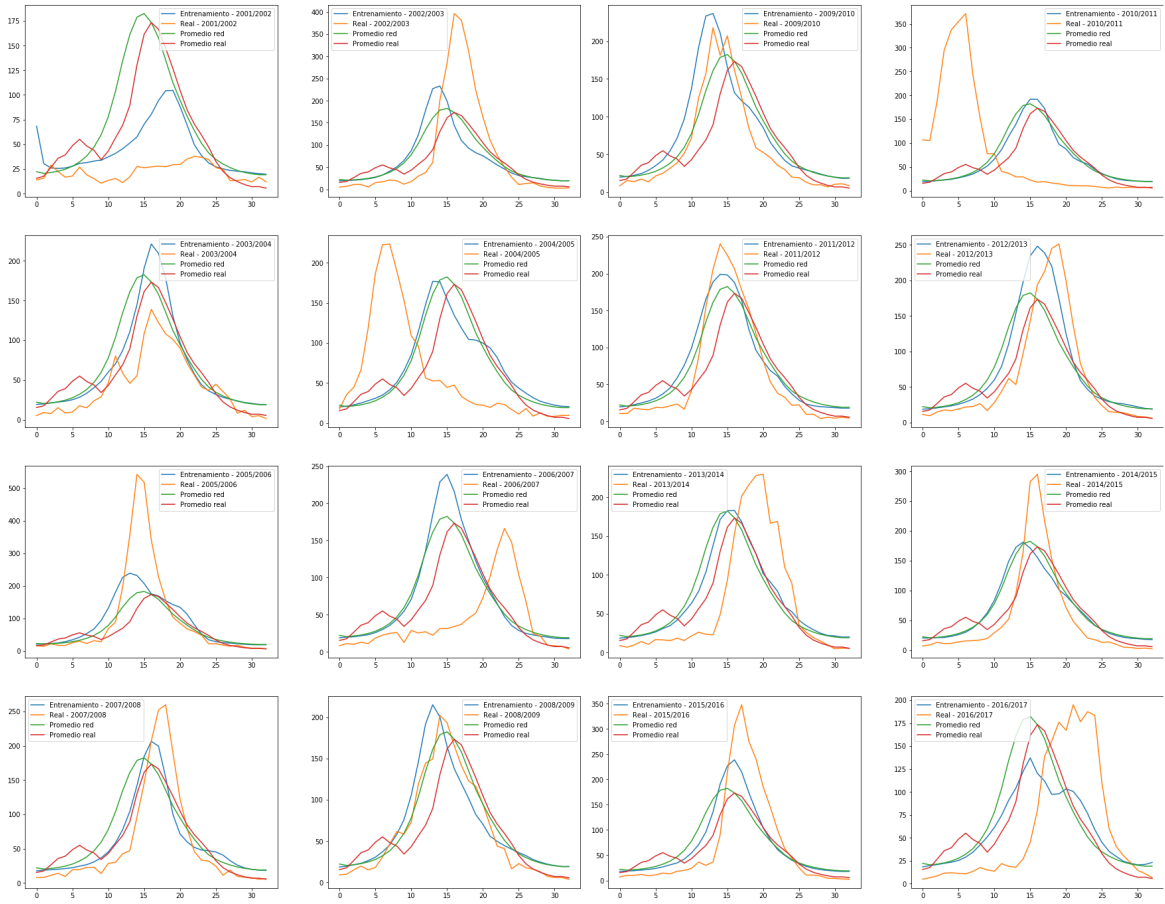


Figura B.2: Resultados de conjunto de entrenamiento. Modelo sin cálculo de $\beta(t)$. Experimento 2

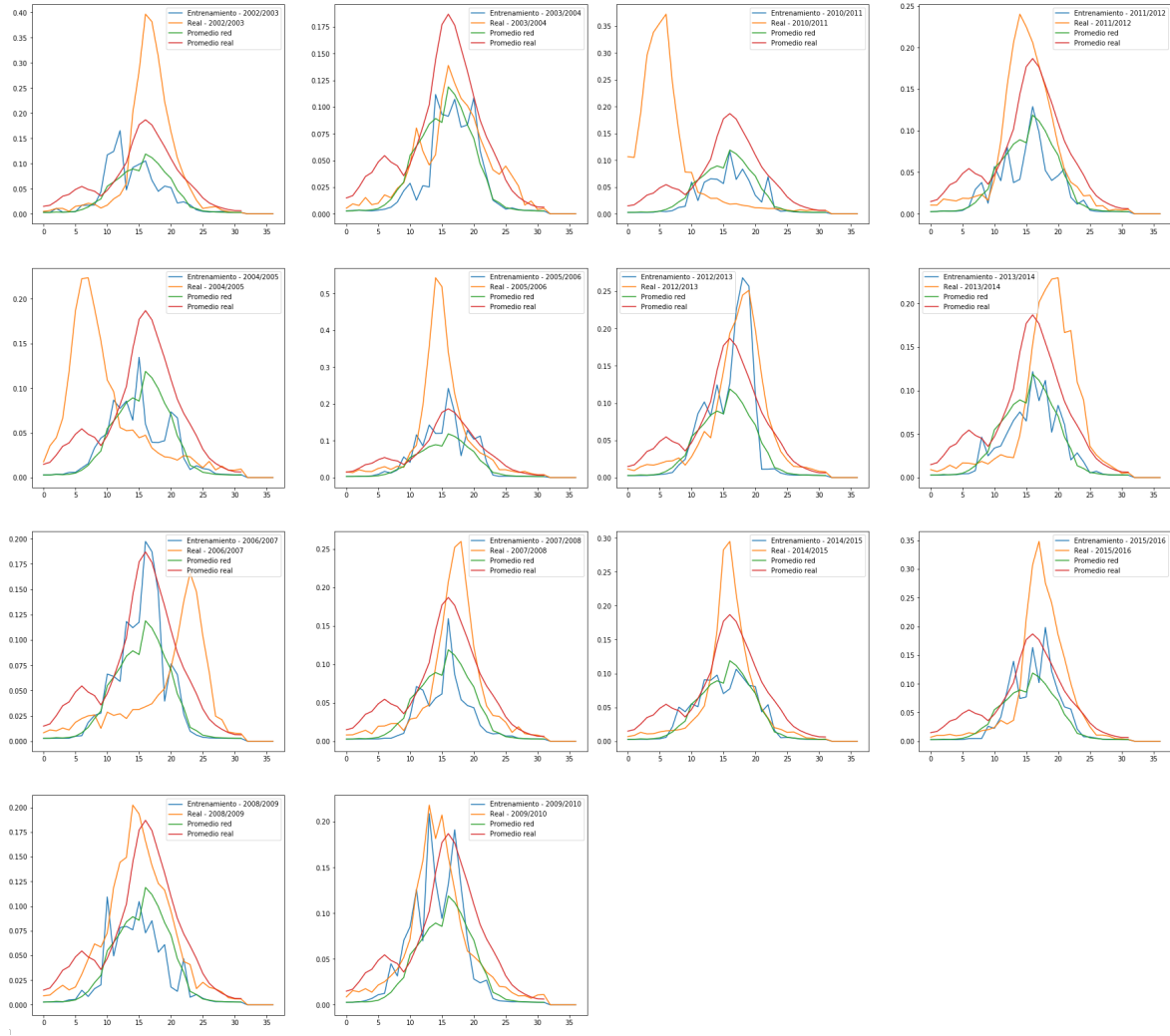


Figura B.3: Resultados de conjunto de entrenamiento. Modelo sin cálculo de $\beta(t)$. Experimento 3

B.2. Resultados modelo predictivo de variable beta

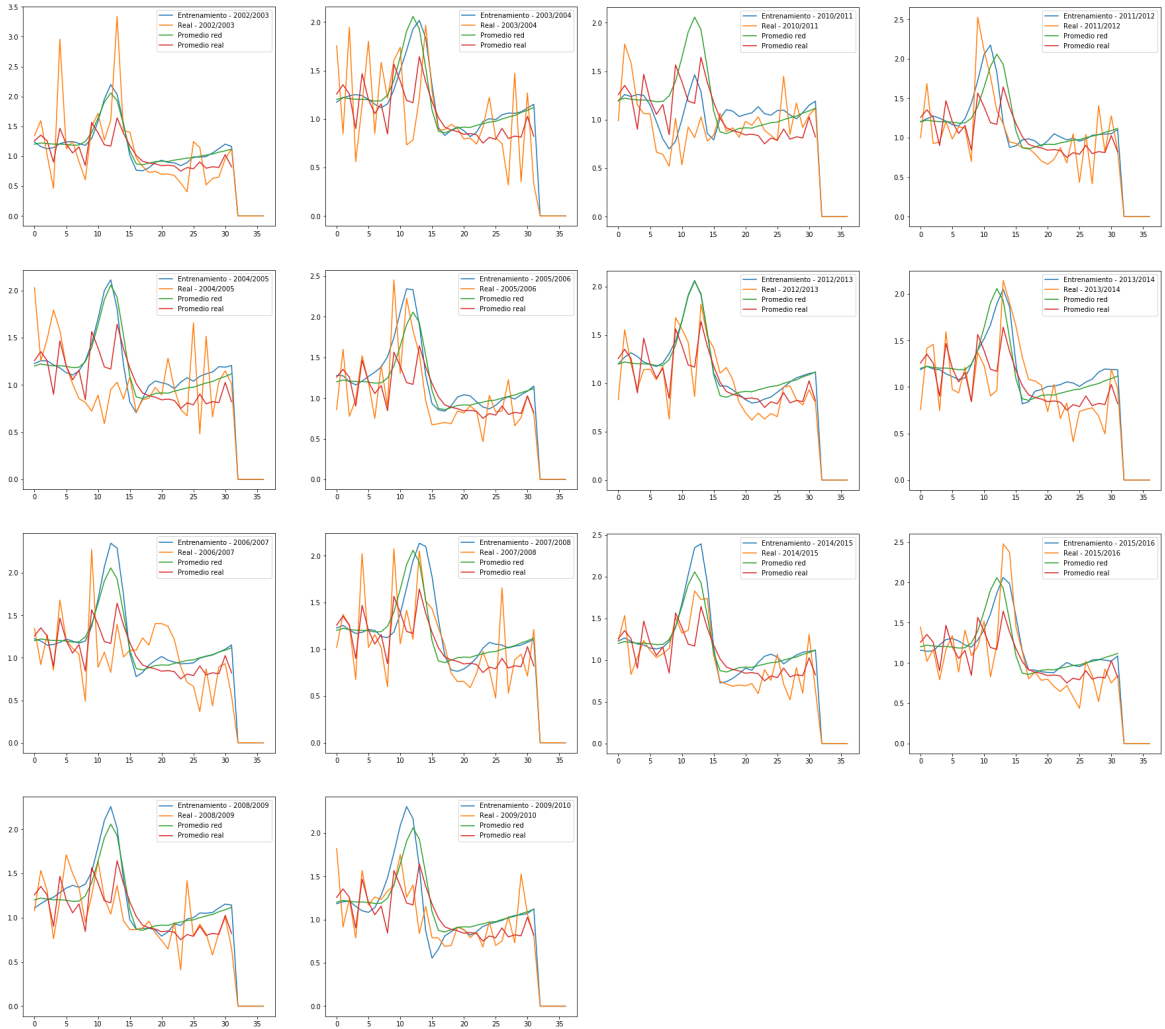


Figura B.4: Resultados de conjunto de entrenamiento. Modelo con cálculo de $\beta(t)$. Experimento 4.

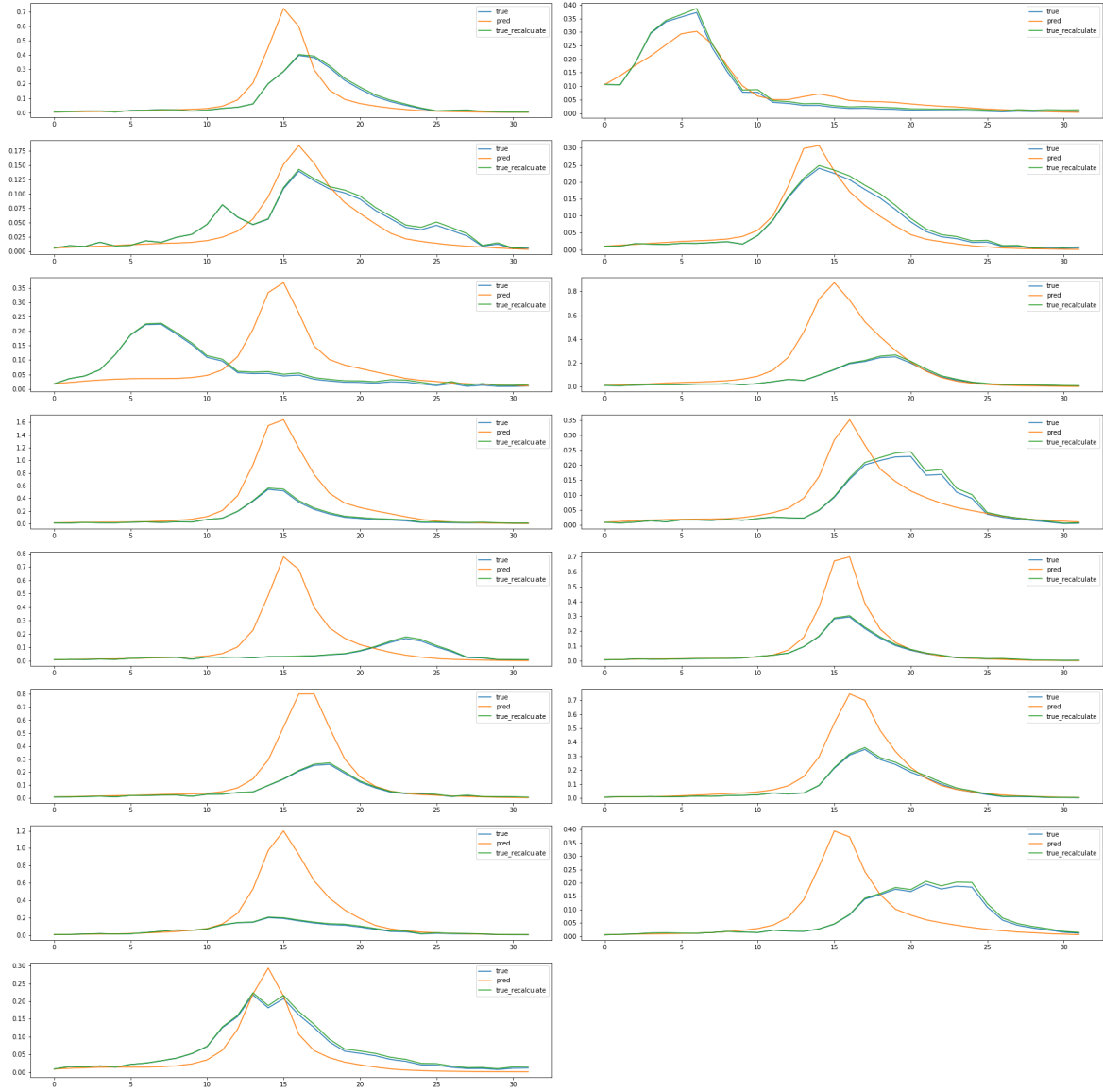


Figura B.5: Resultados de conjunto de entrenamiento. Aplicando $\beta(t)$ calculada. Experimento 4. Leyenda: "true" datos reales, "pred" predicciones de la red y "true_recalculate" datos reproducidos a partir de la obtención de $\beta(t)$ mediante el sistema SIR.

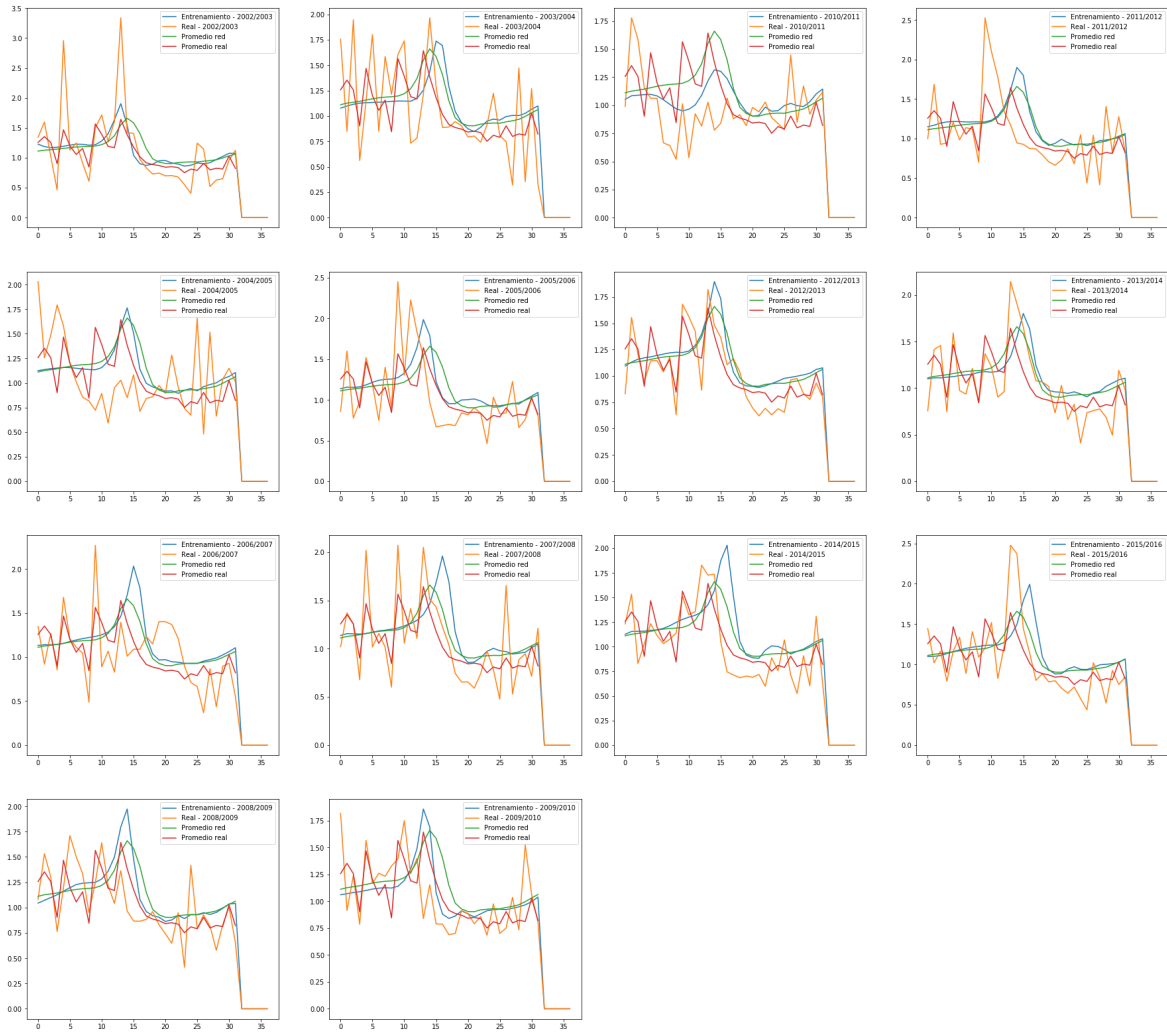


Figura B.6: Resultados de conjunto de entrenamiento. Modelo con cálculo de $\beta(t)$. Experimento 5.

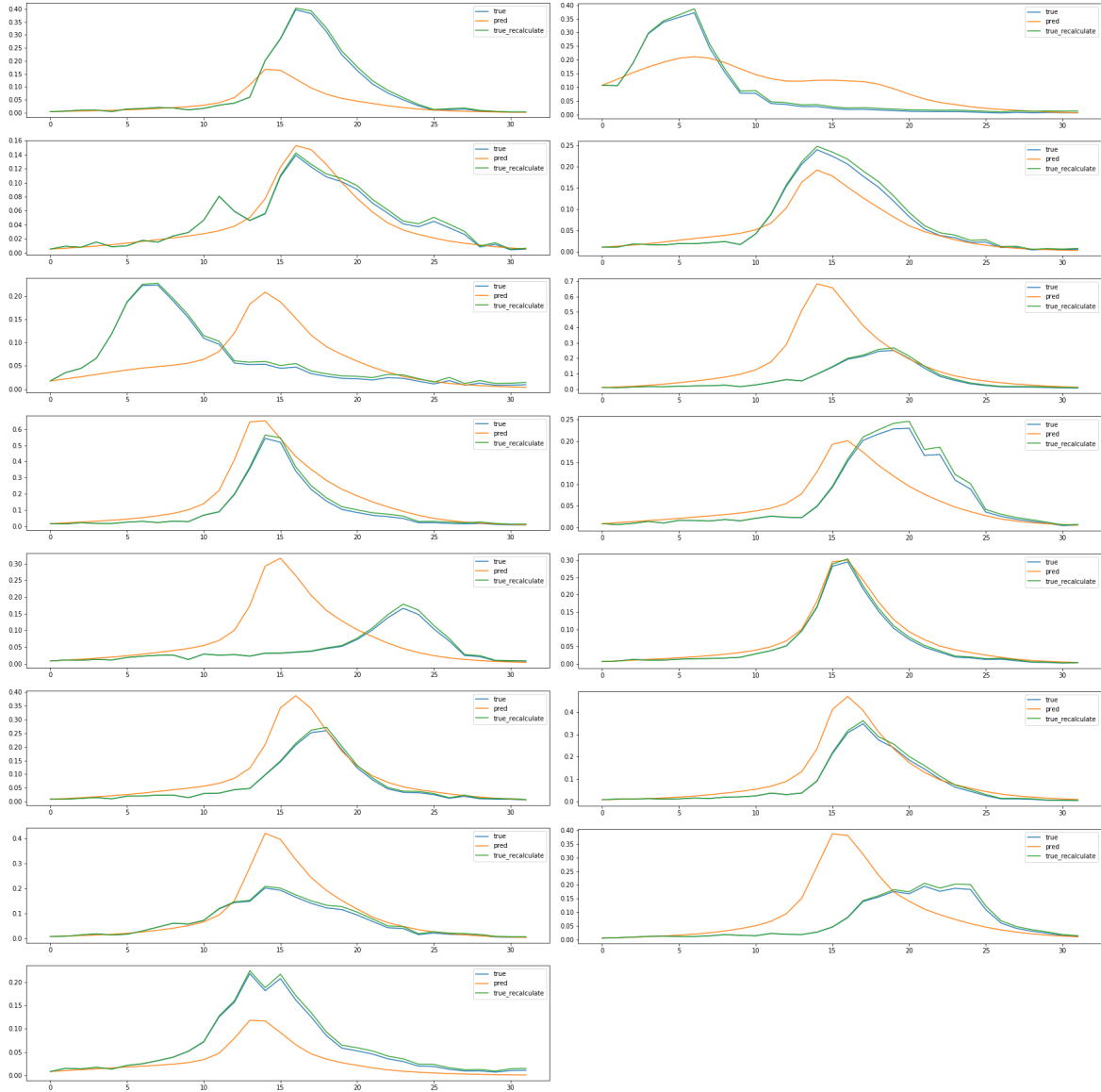


Figura B.7: Resultados de conjunto de entrenamiento. Aplicando $\beta(t)$ calculada. Experimento 5.
 Leyenda: "true" datos reales, "pred" predicciones de la red y "true_recalculate" datos reproducidos a partir de la obtención de $\beta(t)$ mediante el sistema SIR.

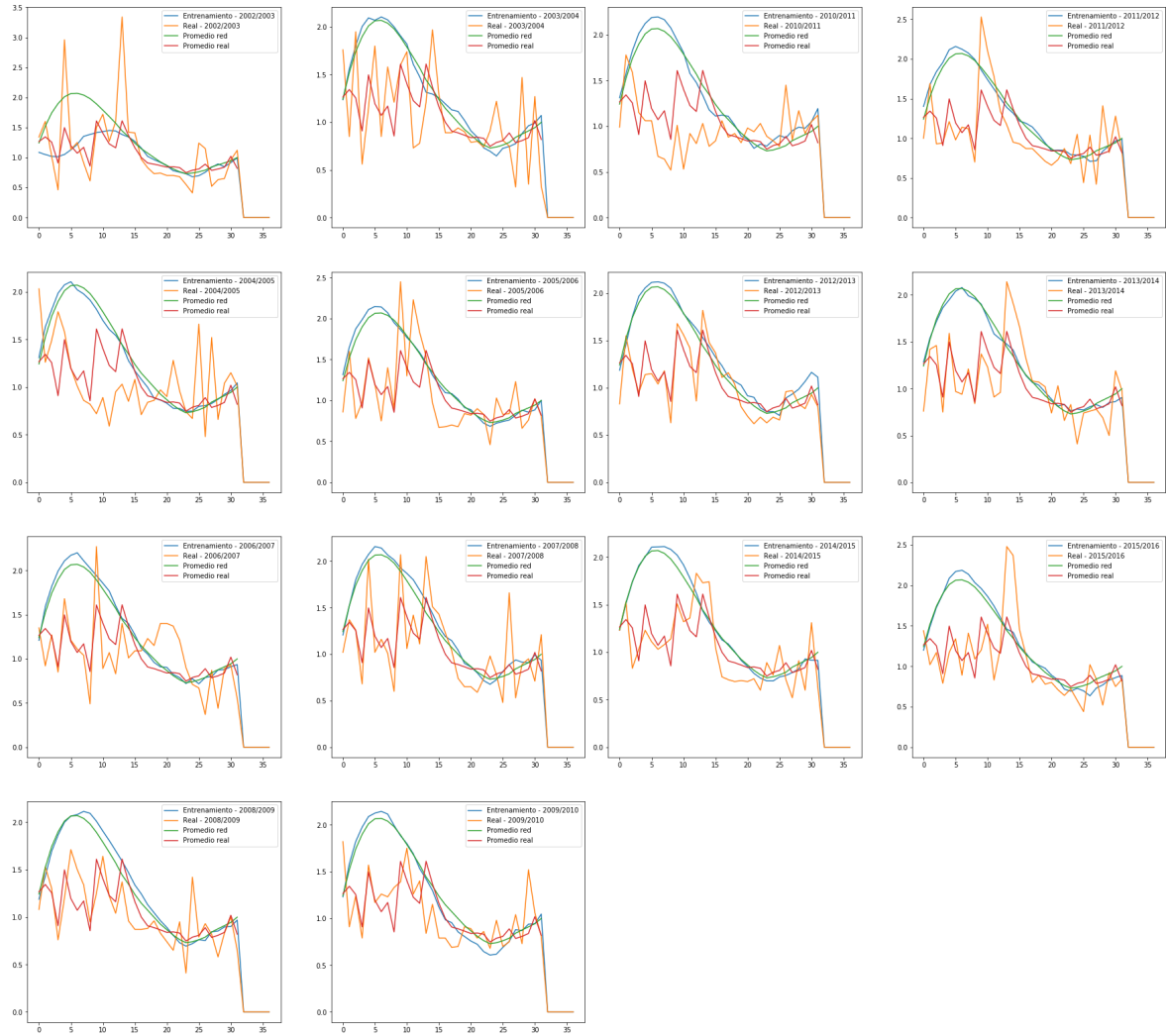


Figura B.8: Resultados de conjunto de entrenamiento. Modelo con cálculo de $\beta(t)$. Experimento 6.

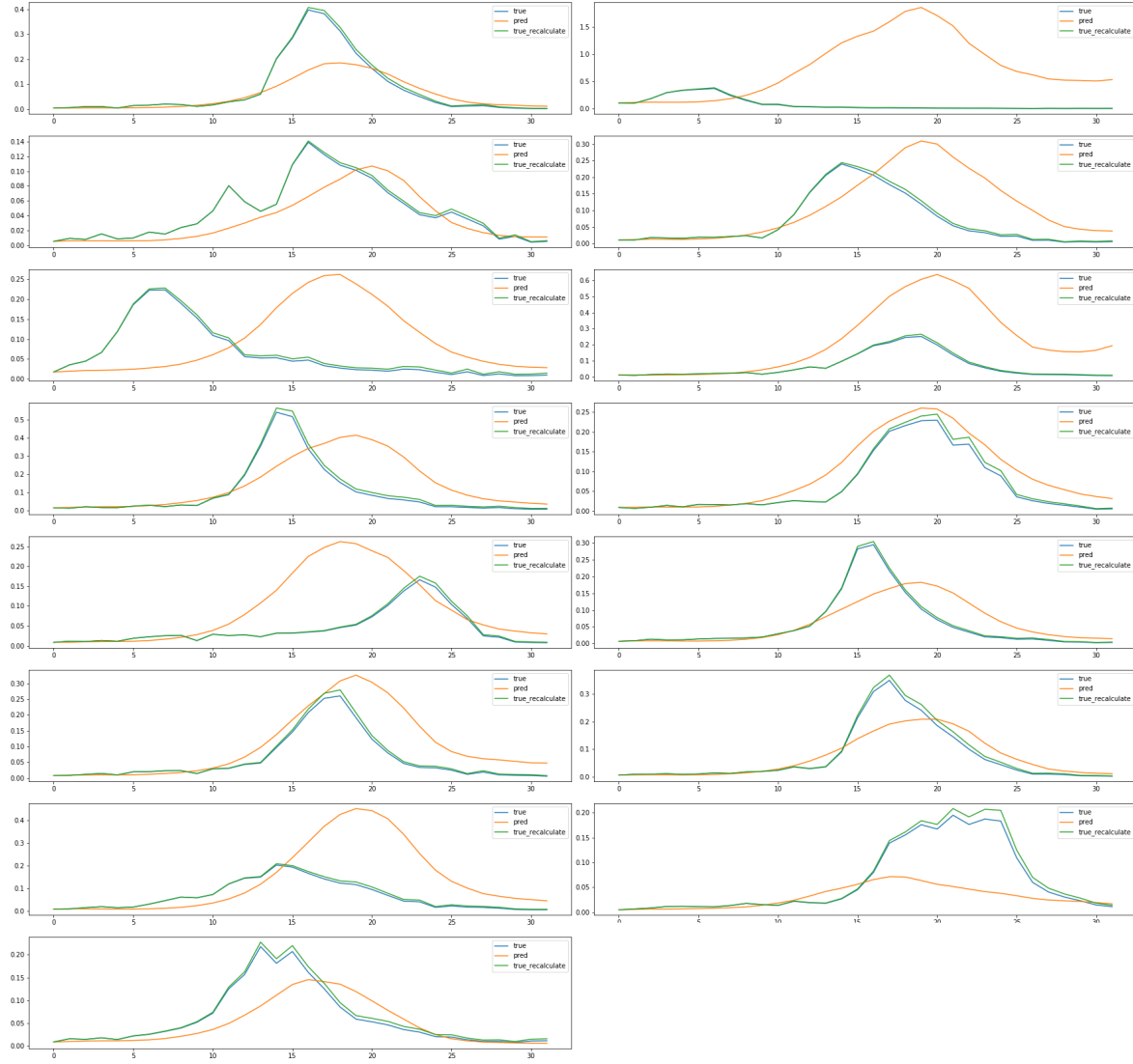


Figura B.9: Resultados de conjunto de entrenamiento. Aplicando $\beta(t)$ calculada. Experimento 6. Leyenda: "true" datos reales, "pred" predicciones de la red y "true_recalculate" datos reproducidos a partir de la obtención de $\beta(t)$ mediante el sistema SIR.